DETERMINATION OF SMOOTHING FOR SPECTRAL-MATRIX ESTIMATION

Walter S. Liggett, Jr.

Raytheon Company

Prepared for:

Office of Naval Research

26 October 1973

UNCLASSIFIED
Security Classification

AD 769 922

DOCUMENT CONTROL DATA - R & D
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Raytheon Company, Submarine Signal Division
Portsmouth, Rhode Island 02871

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
Determination of smoothing for Spectral-Matrix Estimation

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report

5. AUTHOR(S) (First name, middle initial, last name)
Liggett, Walter S., Jr.

6. REPORT DATE
October 26, 1973

7a. TOTAL NO. OF PAGES

7b. NO. OF REFS
32

8a. CONTRACT OR GRANT NO.
N00014- 72- C- 0314

b. PROJECT NO.
NR 042-287

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Statistics and Probability Programs Code 436
Office of Naval Research
Arlington, Virginia 22217

13. ABSTRACT

A statistical procedure for determining the resolution to be used in spectral-matrix estimation is needed for the exploratory analysis of non-stationary multiple time series. Our procedure accomplished this by creating periodograms for a frequency-time interval and then successively dividing this interval into subintervals. At each step, the division is chosen by comparing spectral matrices using the largest-root statistic. This multiple-decision rule is shown to be consistent. Small-sample biases and approximate critical regions for testing stationarity and frequency uniformity are found. Simulation shows that simultaneous analysis of all the series performs better than individual analyses.

53

DD FORM 1 NOV 68 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Multiple Time Series						
Spectral Analysis						
Clustering						
Non-stationarity						
Complex Multivariate Analysis						
Exploratory Data Analysis						

DETERMINATION OF SMOOTHING FOR SPECTRAL MATRIX ESTIMATION

BY

WALTER S. LIGGETT, JR.

October 19, 1973

Research supported by the Office of Naval Research
under Contract N00014-72-C-0314 (NR 042-287)

RAYTHEON COMPANY
SUBMARINE SIGNAL DIVISION
Portsmouth, Rhode Island 02871

DETERMINATION OF SMOOTHING FOR SPECTRAL-MATRIX ESTIMATION

Walter S. Liggett, Jr.

ABSTRACT

A statistical procedure for determining the resolution to be used in spectral-matrix estimation is needed for the exploratory analysis of non-stationary multiple time series. Our procedure accomplishes this by creating periodograms for a frequency-time interval and then successively dividing this interval into subintervals. At each step, the division is chosen by comparing spectral matrices using the largest-root statistic. This multiple-decision rule is shown to be consistent. Small-sample biases and approximate critical regions for testing stationarity and frequency uniformity are found. Simulation shows that simultaneous analysis of all the series performs better than individual analyses.

1. INTRODUCTION

In all spectral analysis, a key problem is choosing the resolution so that the spectral estimates have adequate stability without unacceptable bias. Compared to the case of a single, stationary time series, this problem is significantly more complicated in the case of non-stationary or multiple time series. In particular, the usual techniques for determining the proper frequency or time resolution from the data are not effective for non-stationary multiple time series. In this paper, we analyze a technique suitable for non-stationary multiple time series which we proposed previously [20].

The time series for which our procedure is intended are those with slowly-varying covariance properties and those for which spectral analysis is appropriate. Slowly varying means that in effect the covariance matrix depends only on the time difference within time intervals over which statistical dependence exists. This notion, which justifies the estimation of the spectral matrix as a function of time, has been made more precise by Priestley [25]. The time series for which spectral analysis is appropriate are generally large, have zero mean or the mean removed, and have characteristics of interest that are most naturally expressed in the frequency domain. Such processes have been considered in the statistics literature [11, 25, 28]. Variation of the spectral properties with time is often of interest in applications.

Spectral-matrix estimates are rarely the end result; they usually must be interpreted. Beside graphical interpretation of the spectrum, multiple time series can be analyzed by applying to the spectral matrix techniques useful for interpreting covariance matrices such as principal-component analysis [2, 3, 10, 12, 27].

The bias-stability trade-off for non-stationary multiple time series differs in three important ways from the trade-off for a single stationary time series. Generally, these differences lead to requirements for both more resolution and more

stability in the former case. The first difference is that for non-stationary processes, inadequate time resolution introduces smearing in time that is equivalent to the smearing introduced by inadequate frequency resolution. If the spectral estimate is formed from data in a time interval during which the population spectral matrix changes, the mean of the estimate will be an average of the population spectral matrix over the time interval. To understand the second difference, consider a time-varying linear filter with one input and several outputs. Since there is only one input, the population spectral matrix of the outputs will have rank one. However, a spectral-matrix estimate obtained from an interval in which the relations among the outputs vary with frequency or time will not have rank one. This is the bias that can cause a coherence estimate that should be one to be much less than one. Analogous smearing can occur with more complicated spectral-matrix structures. The third difference is the requirement of the techniques for spectral-matrix interpretation for more stability than is needed for interpretation of the spectrum. The procedures of complex multivariate analysis have biases due to inadequate stability because they form non-linear functions of the spectral-matrix estimates. For example, the usual coherence estimates are biased upward due to inadequate stability. Thus, in coherence estimation, there is a trade-off between two biases which act in opposite directions [30].

The procedures that have been proposed for determining the proper resolution from the data fall into four categories, those based on visual interpretation of data displayed versus frequency and time, those that test for serial correlation, those that determine the proper frequency resolution for stationary time series, and those that test for stationarity. Our procedure differs from each of these. Although we have not done a thorough performance comparison, our procedure seems to have unique capabilities useful in many situations including the exploratory analysis of multiple time series with many components.

A time-honored approach to determining the frequency and time resolution for a single time series is the spectrogram [18, 29]. A spectrogram is created by calculating the periodograms for successive time intervals and then displaying

them as intensity versus frequency and time. As a tool for exploratory analysis, the spectrogram is undoubtedly effective for a single time series. Extension to multiple time series is largely blocked by the inability to simultaneously display the p^2 independent variables (for p time series) of the spectral matrix versus frequency and time.

Our procedure is a generalization of the test for serial correlation in a single stationary time series presented by Durbin [8]. This test compares the sample path of the cumulated periodogram with Kolmogorov-Smirnov limits. Section 3 clarifies this relation.

For stationary time series, the technique of trying different frequency resolutions has often been suggested. One procedure is a graphical comparison of spectra with different resolutions [15, p. 280]. In autoregressive spectral estimation, a more exact procedure is possible for choosing the proper order for the autoregressive scheme and thus the resolution [1, 23]. This latter procedure extends to multiple time series. Except for special types of nonstationarity, these procedures have not been extended to finding the time resolution for non-stationary time series.

Tests for stationarity have been proposed [22, p. 405; 26; 28]. These tests are based on comparison of the results from non-overlapping frequency-time intervals that are large enough to allow stable spectral-matrix estimates to be obtained. Reducing the problem to comparing results that are independent and identically distributed under the null hypothesis has great advantages including simplicity and robustness. For example, Priestley reduces the problem to the analysis of variance [26, 28]. Unfortunately, determining sufficiently large frequency-time intervals a priori is not always possible. When the spectral properties vary in both frequency and time, intervals with the same shape that are sufficiently large and have constant spectral properties may not even exist. Thus, our procedure which is capable of choosing intervals with a wider variety of shapes is sometimes necessary despite the increased complexity. This is especially true in exploratory studies where uncovering the unexpected is desired [7, 31].

In the following sections, we describe our procedure, discuss its statistical properties, and report the results of simulation. In Section 2, we describe the rationale for our procedure and some alternatives which might be useful. Analysis of the dependence of one step of our procedure on the previous ones does not seem possible. Thus, in Section 3, we derive statistical properties of one step in isolation including asymptotic consistency, univariate and approximate joint distributions for our measure of dissimilarity for the case of constant spectral properties, and small-sample biases. From these results, approximate critical regions for testing stationarity and uniformity in frequency can be computed. In Section 4, we demonstrate by simulation that analyzing an entire multiple time series simultaneously performs better than analyzing each component time series individually.

2. CLUSTERING SPECTRAL ESTIMATES

This paper presents a procedure for clustering the spectral-matrix estimates corresponding to different frequencies and times. By Fourier transforming the data in successive time intervals, matrix analogues of (modified) periodogram coefficients can be obtained for a grid of points in frequency and time. These elementary spectral-matrix estimates are approximately independent but not identically distributed when the population spectral matrix varies with frequency or time. Since these estimates are unstable, they must be averaged to make them useful. If the elementary spectral-matrix estimates are nearly unbiased, the problem of obtaining good spectral estimates is reduced to choosing frequency-time intervals in which the population spectral matrix is nearly constant and then smoothing over these intervals. Our procedure chooses rectangular intervals by successive division of a large rectangular interval.

Although Fourier transforming successive non-overlapping time intervals to obtain periodograms is computationally simple and also the case usually treated in justifying the complex-Wishart distribution, the technique is known to have relatively poor spectral windows. An alternative with better statistical properties can be obtained using faders and overlapping time intervals [4; 12, p. 265]. Since a technique that is appropriate to our procedure has not been detailed previously, we describe one option. We assume that the multiple time series has zero mean.

Let the sample of the (discrete) multiple time series be $x(m)$, $m = 1, \ldots, M$, where m indexes time and $x(m)$ is the column vector of the observations from the p component time series at time m. Let $M = (T + 1/2)L$ where T and L are integers so that $(2T - 2)$ intervals with length 2L and 75% overlap can be created. There are three steps in computing the elementary spectral-matrix estimates. First, from the data in each interval the finite Fourier transforms

$$w(f',\ t') = (4\ \Pi\ L)^{-1/2} \sum_{m=1}^{2L} x(m + t'\ L/2) \exp(-i\ \Pi\ f'\ m/L),$$

$$1 \leqslant f' \leqslant L - 1,\ 0 \leqslant t' \leqslant 2T - 3, \tag{2.1}$$

are computed. We exclude the exceptional frequencies 0 and Π. Second, a cosine-bell fader is applied to give

$$\hat{w}(f',\ t') = -\ w(f' - 1,\ t')/4 + w(f',\ t')/2 - w(f' + 1,\ t')/4$$

$$= (16\ \Pi\ L)^{-1/2} \sum_{m=1}^{2L} x(m + t'\ L/2)\ (1 - \cos(\Pi\ m/L)) \tag{2.2}$$

$$\cdot\ \exp(-i\ \Pi\ f'\ m/L),\ 2 \leqslant f' \leqslant L - 2,\ 0 \leqslant t' \leqslant 2T - 3.$$

Finally, the estimates from adjacent frequencies and times are averaged to produce the elementary spectral-matrix estimates,

$$a(f,\ t) = (2/3) \sum_{f'=2f}^{2f+1} \sum_{t'=2t}^{2t+1} \hat{w}(f',\ t')\ \hat{w}^*(f',\ t'),$$

$$1 \leqslant f \leqslant (L - 3)/2,\ 0 \leqslant t \leqslant T - 2, \tag{2.3}$$

where * denotes conjugate transpose. $a(f,\ t)$ is properly scaled to be an estimate of

$$F(\omega_f) = (1/2\ \Pi) \sum_{m=-\infty}^{\infty} R(m)\ e^{-i\ \omega_f\ m},\ \omega_f = \Pi(4f + 1)/2L, \tag{2.4}$$

where $R(m)$ is the covariance matrix.

Comparison of the frequency window with the time window for our elementary spectral-matrix estimates $a(f,\ t)$ shows considerable similarity. The spectral window of $a(f,\ t)$ is given for large L by

$$k_f(\omega - \omega_f) = \frac{2\ \Pi^3\ L \cos^2\left[L(\omega - \omega_f)\right]\left[L^2(\omega - \omega_f)^2 + 9\ \Pi^2/4\right]}{3\left[(L^2(\omega - \omega_f)^2 - 9\ \Pi^2/4)\ (L^2(\omega - \omega_f)^2 - \Pi^2/4)\right]^2},\ 0 < \omega_f < \Pi. \tag{2.5}$$

The mean of a(f, t) when a(f, t) is formed from a stationary time series with spectrum $F(\omega)$ is

$$\int_{-\Pi}^{\Pi} k_f(\omega - \omega_f)\, F(\omega)\, d\omega, \quad 0 < \omega_f < \Pi. \tag{2.6}$$

For non-stationary time series, an analogous temporal window for a(f, t) is given by

$$k_t(m - \tau) = \begin{cases} (12\,\Pi\,L)^{-1} \left[1 - \cos\left(\Pi\left((m - \tau)/L + 5/4\right)\right)\right]^2 \\ \qquad \text{for } -5/4 \leq (m - \tau)/L \leq -3/4 \\ (12\,\Pi\,L)^{-1} \left\{\left[1 - \cos\left(\Pi\left((m - \tau)/L + 5/4\right)\right)\right]^2 \right. \\ \qquad \left. + \left[1 - \cos\left(\Pi\left((m - \tau)/L + 3/4\right)\right)\right]^2\right\} \\ \qquad \text{for } -3/4 \leq (m - \tau)/L \leq 3/4 \\ (12\,\Pi\,L)^{-1} \left[1 - \cos\left(\Pi\left((m - \tau)/L + 3/4\right)\right)\right]^2 \\ \qquad \text{for } 3/4 \leq (m - \tau)/L \leq 5/4 \\ 0 \qquad \text{otherwise.} \end{cases} \tag{2.7}$$

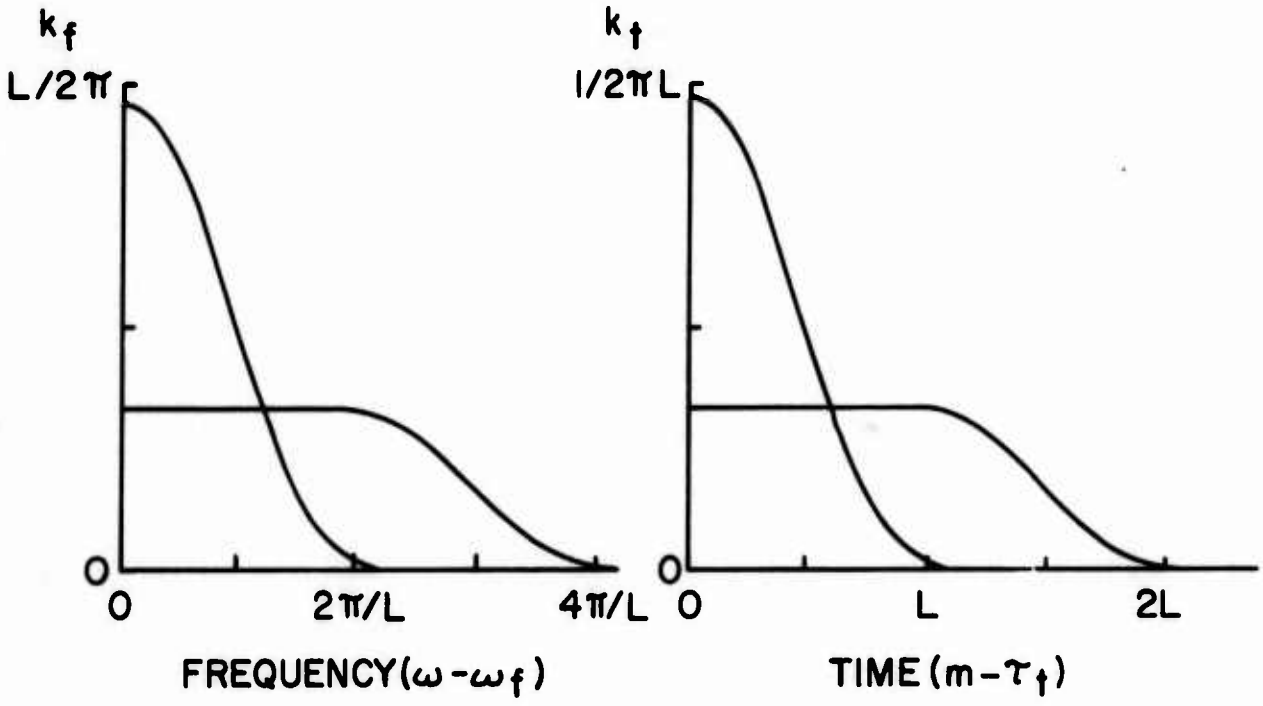
The mean of a(f, t) when a(f, t) is formed from an uncorrelated series with time-varying variances $R_t(0)$ is

$$\sum_{m=-\infty}^{\infty} k_t(m - \tau_t)\, R_m(0), \quad \tau_t = t + L/4\,. \tag{2.8}$$

The windows k_f and k_t are shown in Figure 1 along with the windows for averages of three adjacent elementary spectral matrix estimates. The sidelobes on the frequency window are very small. Within the averaging interval, the windows are nearly flat.

Our procedure is based on the premise that the a(f, t) are independent. This is only approximately true for adjacent estimates as can be seen from the moments of a(f, t) for unit-power white noise ($R_t(\tau) = \delta(\tau)$). For large L, we have

FIGURE I. FREQUENCY AND TIME WINDOWS



$$
\begin{aligned}
&E\, a(f,\ t) = (2\ \Pi)^{-1} \\
&\text{var } a(f,\ t) = (2\ \Pi)^{-2}\ (0.527) \\
&\text{cov } (a(f,\ t),\ a(f+1,\ t)) = (2\ \Pi)^{-2}\ (0.095) \\
&\text{cov } (a(f,\ t),\ a(f,\ t+1)) = (2\ \Pi)^{-2}\ (0.095).
\end{aligned}
\tag{2.9}
$$

If a spectral estimate is formed by averaging $a(f,\ t)$ over $f_L \leq f \leq f_U$, $t_L \leq t \leq t_U$ where $f_U - f_L$ and $t_U - t_L$ are both large, then the ratio of the variance of this estimate to its mean will be nearly that expected under statistical independence.

In the remainder of this paper, we proceed as though the $a(f,\ t)$ are independent and distributed as $\xi\xi^*$, where ξ is complex-Gaussian distributed with zero mean and complex covariance matrix $F(f,\ t)$ ($F(f,\ t)$ is the spectral matrix at frequency f and time t). The primary consequence of this is that sums of $a(f,\ t)$ will be treated as complex-Wishart distributed. Thus, the fact that $a(f,\ t)$ does not have rank one (except when $p = 1$) will cause no apparent difficulty. The use of the complex-Wishart distribution as the distribution of a spectral-matrix estimate is almost always an approximation. The approximation is asymptotically valid under various circumstances including cases where the input time series is not Gaussian [12, p. 220; 32]. There is no asymptotic result which completely justifies our approximation.

Our procedure for smoothing the elementary spectral matrix estimates $a(f,\ t)$ is a divisive clustering algorithm that creates clusters connected in the frequency-time plane. Connected clusters are required because we assume nothing about the population spectral matrix $F(f,\ t)$ except smoothness in $(f,\ t)$. Thus there is no prior knowledge implying that two points in frequency-time have the same spectral matrix without implying that the points in between do also. Before describing our procedure in detail, we consider alternatives to our divisive algorithm and our dissimilarity measure.

Our algorithm starts with a large rectangular interval in the frequency-time plane and partitions this into rectangular subintervals. At each step, one rectangular subinterval is divided in two. There are several considerations relevant to adopting this scheme. One advantage of a hierarchical scheme is the possibility that for most inputs the gross behavior of each step will be simple enough to be understood even though analysis of the statistical dependence of one step on the previous ones is not tractable. More specifically, our scheme which selects one division at each step is simpler than a procedure that selects several divisions at each step. Another advantage is that hierarchical schemes require fewer computations than the most general schemes. The advantage of a divisive scheme is that the first steps are based on large blocks of data for which the small-sample biases are less important. Since the small-sample biases of most multivariate criteria are hard to evaluate, these biases should be avoided as long as possible. Another aspect of this is that for the first step the problem of statistical dependence on preceding steps does not apply so that the first step can be analyzed more exactly. Thus, the first step can be interpreted for its own significance such as testing for the uniformity of the population spectral matrix within the original interval. Some alternative procedures can select partitions of the original interval that our procedure cannot select. Compared to these alternatives, our restricted range of partitions is an advantage because the effective number of choices is not too great for the available data and because the resulting partition is simple enough for visual interpretation. The disadvantage is that none of the possible partitions may fit the data very well.

In our procedure, we select each division by comparing the spectral matrices created from the subintervals on either side of a potential division. The criterion is the largest-root statistic adjusted for the variation of the size of the two subintervals. At least four ways to compare spectral matrices can be adapted from tests for comparing covariance matrices [24]. For differences caused by a rank-one component, the largest-root procedure is the maximum likelihood procedure. Another advantage of our choice is that since spectral estimates on both sides of a

potential division do not have to have full rank, a small interval with a high spectral level can be isolated.

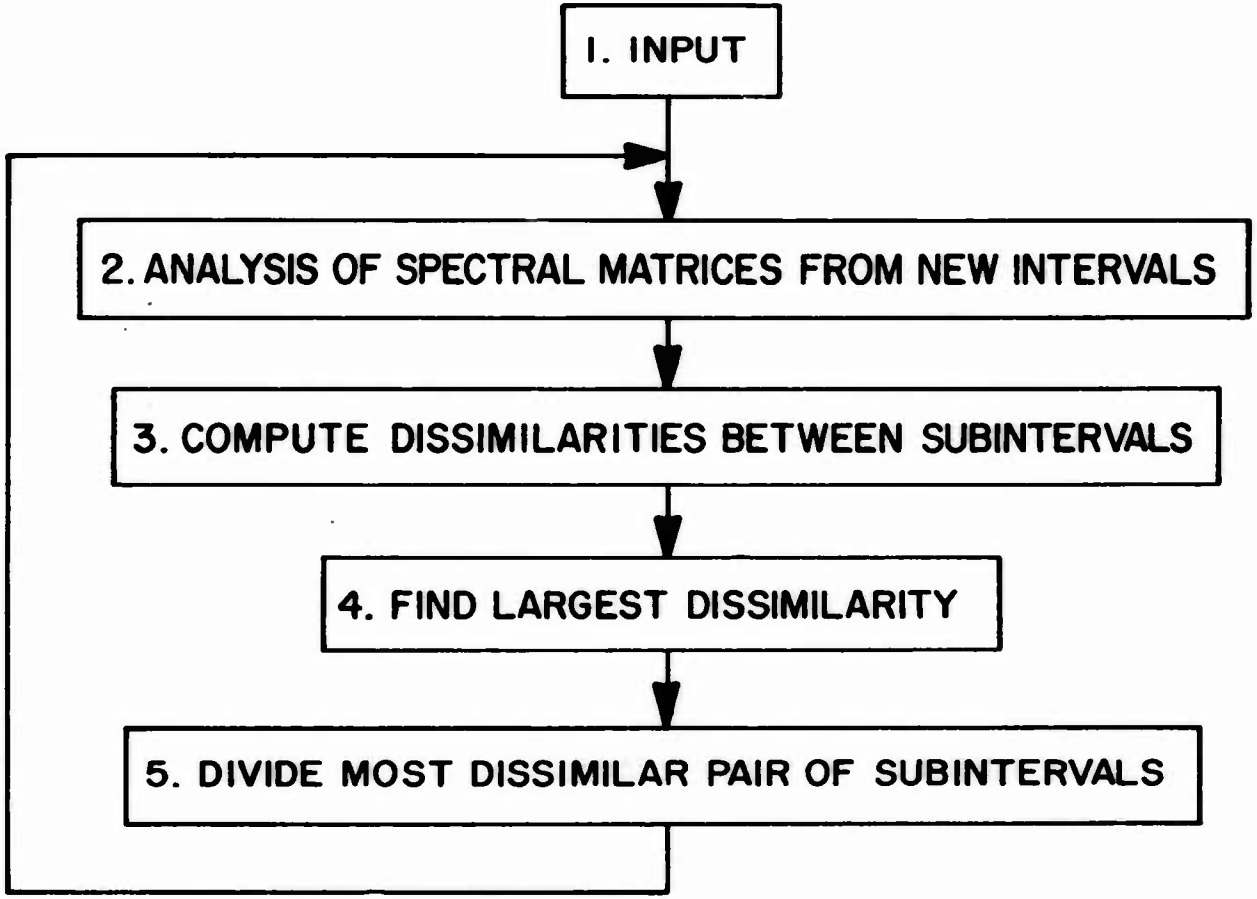
As mentioned above, one manifestation of bias in spectral-matrix estimation is an estimate with more significant principal components than the population spectral matrix due to smoothing over an interval that is too large. This suggests that intervals might be selected so that the number of significant components is minimized in some sense. Algorithms like this have been suggested for real multivariate data [9, p. 289; 13]. These algorithms apply principal-component analysis to subsets of the data set to discover whether the data lies close to a lower-dimensional subspace.

We complete the description of our procedure by discussing the flow diagram shown in Figure 2. The inputs are a(f, t) for (f, t) in a large rectangular interval. This interval is chosen on the basis of prior knowledge to be so large that smoothing over a region that extends from its center to beyond its borders will incur unacceptable bias. In the second block, the spectral matrix from each interval in the sequence of partitions is analyzed using principal components or techniques appropriate to the application. This step does not influence the sequence of partitions, but it does provide outputs helpful in choosing the right member of the sequence. The third block computes the measure of dissimilarity for all potential divisions. The measure of dissimiliarity for a potential division is the largest of two one-sided measures that correspond to the two potential new intervals. We formulate our procedure in terms of these one-sided measures. To specify these one-sided measures, let

$$S_0 = \{(f, t) \mid f_L \leq f \leq f_U,\ t_L \leq t \leq t_U\} \tag{2.10}$$

be an interval already created by the procedure. The new intervals that might be created by dividing S_0 are those intervals S for which either S or $S_0 - S$ is given by $S = \{(f, t) \mid f_L \leq f \leq f_B,\ t_L \leq t \leq t_U\}$ where $f_L \leq f_B < f_U$ or $S = \{(f, t) \mid f_L \leq f \leq f_U,\ t_L \leq t \leq t_B\}$ where $t_L \leq t_B < t_U$. Let N_0 be the number of points in S_0, let N be the number of points in S, and let

FIGURE 2. DIAGRAM OF THE PROCEDURE



$$A_0 = \sum_{(f,\, t) \in S_0} a\,(f,\, t)\ ,\ A = \sum_{(f,\, t) \in S} a\,(f,\, t). \tag{2.11}$$

The one-sided measure of dissimilarity is

$$d_e(S) = \begin{cases} d(S) & \text{if } N \leq N_0 - 4p \\ 0 & \text{if } N > N_0 - 4p \text{ or } N_0 < 8p, \end{cases} \tag{2.12}$$

where

$$d(S) = \begin{cases} \left\{ 2N \log \left[N/(N_0 \psi) \right] + 2\,(N_0 - N) \log \left[(N_0 - N)/(N_0(1 - \psi)) \right] \right\}^{1/2} & \text{if } \psi > N/N_0 \\ 0 & \text{otherwise} \end{cases} \tag{2.13}$$

and ψ is the largest eigenvalue of $A_0^{-1} A$. As discussed in the next section, the exclusion of values of d(S) indicated in (2.12) is intended to compensate for a small-sample bias. In block 4, the procedure finds the largest d_e among those computed from all permitted subintervals of all the intervals previously created. In block 5, the subinterval corresponding to the largest d_e and the subinterval containing the remainder of S_0 become the new intervals in the partition. Equation (2.12) excludes divisions of intervals with less than 8p points. Since S_0 might be divided into S and $S_0 - S$ in two ways, $d_e(S)$ is largest or $d_e(S_0 - S)$ is largest, (2.12) does not exclude any division of any interval with $N_0 \geq 8p$.

Starting with the original interval, our procedure creates a sequence of successively finer partitions until the largest d_e is zero. There are three guides useful in determining which member of this sequence to use in creating the final spectral-matrix estimate. As discussed in the next section, one guide is the sequence of values of the maximum dissimilarity. In particular, regions for testing the hypothesis that the first division corresponds to a difference in the population spectral matrix versus the hypothesis that the population spectral matrix is uniform are provided. Another guide is the outputs at block 2. These outputs can be used

to judge whether or not a division reduces the number of significant principal components. Finally, the configuration of the intervals in each partition can be used along with prior knowledge to guide the choice of partition.

In many applications of spectral analysis, one objective is the discovery of narrow frequency bands with high spectral levels. If such a band were in the middle of an interval, the set of potential divisions in our procedure might not be most sensitive. For example, comparison of the spectral matrices from the subintervals $S_N = \{(f, t) \mid f_L < f_1 \leq f \leq f_2 < f_U,\ t_L \leq t \leq t_U\}$ with the spectral matrices from $S_0 - S_N$ might be more likely to isolate a narrowband. Such comparisons, which would result in two divisions instead of one, can be included in our procedure. Many of the results in the next section apply to such an extension.

3. ANALYSIS OF THE CRITERION

The core of our procedure is the selection of the two subintervals to which correspond the most dissimilar spectral matrices. This section is devoted to the analysis of this selection for a given interval. The results are rigorously applicable to the first division of the original interval. They can be used for guidance in interpreting further divisions but not with the same rigor since the statistical dependence of one step on its predecessors has not been included. Because the number of alternatives at each step is small compared to the amount of data, the guidance provided seems applicable at least for the first few steps.

3.1 Maximum Likelihood

Our algorithm for selecting the most dissimilar pair of spectral matrices is the maximum likelihood procedure for finding two subintervals whose spectral matrices differ by a rank-one component. Confining the set of possible divisions to pairs of rectangular subintervals as our procedure does, we prove asymptotic consistency. Consistency shows that if the interval is large enough, the bias does not dominate the real spectral properties. Consistency is one indication that the number of choices is not too large for the amount of data.

Theorem 1: Let S_0 be an interval in the frequency-time plane with N_0 points, where $N_0 > p$. Let $\mathcal{S}$ be a set of subsets of S_0 with members that are neither empty nor have more than $N_0 - p$ points. Let the $a(f, t)$ be independent and distributed as $\xi\xi^*$ where ξ is complex Gaussian with zero mean and complex covariance matrix $F_0 + uu^*$ for $(f, t) \in \tilde{S}$ and F_0 for $(f, t) \in S_0 - \tilde{S}$. Let $\tilde{S} \in \mathcal{S}$ and F_0 be positive definite. The maximum likelihood estimate of $\tilde{S}$ is the $\tilde{S} \in \mathcal{S}$ that maximizes $d(S)$, which is given by (2.13).

Proof: Beside S, the unknown parameters to be maximized over are the matrix F_0 and the vector u. The likelihood function is

$$L = \Pi^{-pN_0} \left[\det (F_0 + uu^*)\right]^{-N} \left[\det F_0\right]^{-(N_0 - N)}$$

$$\cdot \exp\left\{- \operatorname{tr}\left[(F_0 + uu^*)^{-1} A + F_0^{-1} (A_0 - A)\right]\right\}, \tag{3.1}$$

where A_0 and A are given by (2.11). With probability one, the rank of A is greater than zero and the rank of $A_0 - A$ is p. Maximizing over u with F_0 fixed gives

$$\max_u L = \Pi^{-pN_0} \left(\det F_0\right)^{-N_0} (\alpha/N)^{-N} \exp\left\{- \operatorname{tr} F_0^{-1} A_0 + \alpha - N\right\}, \tag{3.2}$$

where

$$\alpha = \max\left\{N, \max_{w^*w=1} w^* F_0^{-1/2} A F_0^{-1/2} w\right\}. \tag{3.3}$$

Consider first the case where A is non-singular and let λ_j and ϕ_j be the eigenvalues and eigenvectors of $A^{1/2} F_0^{-1} A^{1/2}$ in decreasing order. Equation (3.2) becomes

$$\max_u L = \Pi^{-pN_0} \left(\det A\right)^{-N_0} \prod_{j=1}^{p} \lambda_j^{N_0} (\alpha/N)^{-N}$$

$$\cdot \exp\left\{- \sum_{j=1}^{p} \lambda_j \beta_j + \alpha - N\right\}, \tag{3.4}$$

where

$$\beta_j = \phi_j^* A^{-1/2} A_0 A^{-1/2} \phi_j, \tag{3.5}$$

$$\alpha = \max\left\{N, \lambda_1\right\}.$$

Using the result

$$g(\beta_1) = \max_{\lambda_1} \left[\lambda_1^{N_0} \alpha^{-N} \exp\left\{ -\lambda_1 \beta_1 + \alpha \right\} \right] = \begin{cases} \left(\frac{N_0 - N}{\beta_1 - 1} \right)^{N_0 - N} e^{-(N_0 - N)} & \text{if } 1 < \beta_1 \leq \frac{N_0}{N} \\ \left(\frac{N_0}{\beta_1} \right)^{N_0} N^{-N} e^{-(N_0 - N)} & \text{if } \frac{N_0}{N} < \beta_1, \end{cases} \tag{3.6}$$

we can maximize over $\lambda_1, \lambda_2, \ldots, \lambda_p$. To maximize over $\phi_1, \phi_2, \ldots \phi_p$, we use the result

$$\left[\prod_{j=1}^{p} \alpha_j^{-N_0} \right] \left[\alpha_1^{N_0} g(\alpha_1) \right] \leq \max_{\phi_1, \ldots, \phi_p} \left[\prod_{j=1}^{p} \beta_j^{-N_0} \right] \left[\beta_1^{N_0} g(\beta_1) \right]$$

$$\leq \max_{\phi_1, \ldots, \phi_p} \left[\prod_{j=1}^{p} \beta_j^{-N_0} \right] \max_{\phi_1} \left[\beta_1^{N_0} g(\beta_1) \right] \tag{3.7}$$

$$= \left[\det \left(A^{-1/2} A_0 A^{-1/2} \right) \right]^{-N_0} \alpha_1^{N_0} g(\alpha_1),$$

where $\alpha_1, \alpha_2, \ldots, \alpha_p$ are the eigenvalues of $A^{-1/2} A_0 A^{-1/2}$ in increasing order. The maximization of $\left[\prod \beta_j^{-N_0} \right]$ is a consequence of the solution to the maximization of $(\det B)^{N_0} \exp\left(- \operatorname{tr} B A^{-1/2} A_0 A^{-1/2} \right)$ [10]. Thus, we obtain an expression for the maximum of L over F_0 and u that is a strictly increasing function of d(S). If A is singular, the proof must be modified by initially replacing A by $A + \epsilon I$ and then letting ϵ go to zero after maximization.

In our procedure, the number of possible divisions increases at least as fast as the square root of the number of data points. In our interpretation of consistency,

we account for this increase. Thus, our theorem and proof require some modification of Wald's results [5, p. 54]. The data sequence on which our theorem is based is composed of the a(f, t) for a sequence of nested rectangles in the frequency-time plane,

$$S_0^{(n)} = \left\{ (f,\ t) \;\middle|\; f_L^{(n)} \leq f \leq f_U^{(n)},\ t_L^{(n)} \leq t \leq t_U^{(n)} \right\}. \tag{3.8}$$

We require that as $n \to \infty$, $(f_U^{(n)} - f_L^{(n)})/(t_U^{(n)} - t_L^{(n)}) \to \beta$ where $0 < \beta < \infty$. The parameter space that contains the true parameter values and over which the maximum likelihood estimates are formed is given by $F_0 \in \mathcal{F}$, $F_0 + uu^* \in \mathcal{F}$, $u^*u > 0$, and $S \in \mathcal{S}^{(n)}$, where $\mathcal{F}$ is the set of Hermitian matrices F whose eigenvalues lie between ϵ ($\epsilon > 0$) and ∞ and where $\mathcal{S}^{(n)}$ is the set of subsets S that along with $S_0^{(n)}$ - S are non-empty and rectangular. $\mathcal{S}^{(n)}$ is the set of subsets incorporated in our procedure. The true parameter values are $\tilde{F}_0$, $\tilde{u}$, and $\tilde{S}^{(n)}$. Without loss of generality, we let

$$\tilde{S}^{(n)} = \left\{ (f,\ t) \;\middle|\; f_L^{(n)} \leq f \leq f_B < f_U^{(n)},\ t_L^{(n)} \leq t \leq t_U^{(n)} \right\}, \tag{3.9}$$

where as n increases f_B remains fixed and neither $f_B - f_L^{(n)}$ nor $f_U^{(n)} - f_B$ decrease. Let $\tilde{N}$ denote the number of points in $\tilde{S}^{(n)}$.

Theorem 2: Let $\mathcal{N}_0$ be a neighborhood of $\tilde{F}_0$ and $\mathcal{N}_1$ be a neighborhood of $\tilde{F}_1 = \tilde{F}_0 + \tilde{u}\tilde{u}^*$, where $\mathcal{N}_0$ and $\mathcal{N}_1$ are chosen so that for all $F_0 \in \mathcal{N}_0$ and $F_1 \in \mathcal{N}_1$ the largest eigenvalue of $F_0^{-1} F_1$ is greater than and bounded away from one. Let $\hat{S}^{(n)}$, $\hat{F}_0^{(n)}$, and $\hat{u}^{(n)}$ be the sequence of maximum likelihood estimates over the parameter space defined above. In the sequel, we drop the dependence on n from our symbols for the estimates. Then, when the a(f, t) are distributed as above with parameters $\tilde{S}^{(n)}$, $\tilde{F}_0$, and $\tilde{u}$, we have

$$\text{Prob}\left\{ \left[\hat{S} \neq \tilde{S}^{(n)} \right] \cup \left[\hat{F}_0 \notin \mathcal{N}_0 \right] \cup \left[\hat{F}_0 + \hat{u}\hat{u}^* \notin \mathcal{N}_1 \right] \right\} \to 0 \text{ as } n \to \infty. \tag{3.10}$$

Proof: This is an extension of Chernoff's proof [5]. We show that the probability of the event $\left[\hat{\hat{S}} \neq \tilde{S}^{(n)}\right] \cup \left[\hat{\hat{F}}_0 \notin \mathcal{U}_0\right] \cup \left[\hat{\hat{F}}_0 + \hat{\hat{u}}\hat{\hat{u}}^* \notin \mathcal{U}_1\right]$ happening infinitely often is zero. We divide the sequence of occurrences of this event into a finite number of subsequences and then prove that the probability of any subsequence occurring infinitely often is zero. First, we divide the event into the cases: $\hat{\hat{S}}$ contains the lowest frequencies, $\hat{\hat{S}}$ contains the highest frequencies, $\hat{\hat{S}}$ contains lowest times, and $\hat{\hat{S}}$ contains the highest times. Each of these cases must be further divided.

In order to do this, we extend the parameter space to a compact set. We add the points for which u*u = 0 and create the compact set $\mathcal{F}^*$ out of $\mathcal{F}$ by adding the point at which $\max_{w^*w=1} w^* F w = \infty$. As in [5], we create a finite covering of $\mathcal{F}^*$ with members $\mathcal{U}$ such that for both j = 0 and j = 1 if $\mathcal{U}_j \not\supset \mathcal{U}$ then

$$\alpha_j(\mathcal{U}) = E\left\{\inf_{F \in \mathcal{U}} \log\left[\frac{(\det \tilde{F}_j)^{-1} \exp\left\{-\operatorname{tr} \tilde{F}_j^{-1} a(f, t)\right\}}{(\det F)^{-1} \exp\left\{-\operatorname{tr} F^{-1} a(f, t)\right\}}\right]\right\} > 0, \qquad (3.11)$$

where the expectation is with respect to the distribution for which $\tilde{F}_j$ is the spectral matrix. We introduce the notation

$$Q(S, \tilde{F}_j, F) = \sum_{(f, t) \in S} \log\left[\frac{(\det \tilde{F}_j)^{-1} \exp\left\{-\operatorname{tr} \tilde{F}_j^{-1} a(f, t)\right\}}{(\det F)^{-1} \exp\left\{-\operatorname{tr} F^{-1} a(f, t)\right\}}\right]. \qquad (3.12)$$

Consider the events for which $\hat{\hat{S}} \supset \tilde{S}^{(n)}$ and $\hat{\hat{F}}_0 + \hat{\hat{u}}\hat{\hat{u}} \notin \mathcal{U}_1$. For some $\mathcal{U}$, we have

$$\begin{aligned} 0 &\geq \inf_{S \supset \tilde{S}^{(n)}} \inf_{F_0} \inf_{F_1 \in \mathcal{U}} \Big[Q(\tilde{S}^{(n)}, \tilde{F}_1, F_1) \\ &+ Q\left(S \cap (S_0^{(n)} - \tilde{S}^{(n)}), \tilde{F}_0, F_1\right) + Q\left(S_0^{(n)} - S, \tilde{F}_0, F_0\right)\Big] \qquad (3.13)\ (\text{cont}) \\ &\geq \inf_{F_1 \in \mathcal{U}} Q(\tilde{S}^{(n)}, \tilde{F}_1, F_1) \end{aligned}$$

$$+ \inf_S \inf_{F_1 \in \mathcal{F}^*} Q(S \cap (S_0^{(n)} - \tilde{S}^{(n)}),\ \tilde{F}_0,\ F_1) \qquad (3.13)$$

$$+ \inf_S \inf_{F_0 \in \mathcal{F}^*} Q(S_0^{(n)} - S,\ \tilde{F}_0,\ F_0). \qquad \text{(cont)}$$

By the law of large numbers, we have as $n \to \infty$

$$\tilde{N}^{-1} \inf_{F_1 \in \Pi} Q(\tilde{S}^{(n)},\ \tilde{F}_1,\ F_1) \to \alpha_1(\Pi) \text{ a.s.} \qquad (3.14)$$

In the second term on the right side of (3.13), we perform the minimization over F_1 to obtain

$$\inf_S \inf_{F_1 \in \mathcal{F}^*} Q(S \cap (S_0^{(n)} - \tilde{S}^{(n)}),\ \tilde{F}_0,\ F_1) \qquad (3.15)$$

$$= \inf_S \log \left\{ \left[\det (G/N_G) \right]^{N_G} \exp \left[- \operatorname{tr} (G - N_G I) \right] \right\},$$

where

$$G = \tilde{F}_0^{-1} \sum_{(f,\ t) \in S_G} a(f,\ t),\ S_G = S \cap (S_0^{(n)} - \tilde{S}^{(n)}), \qquad (3.16)$$

and N_G is the number of points in S_G. We show that $\tilde{N}^{-1}$ times the right side of (3.15) goes to zero almost surely. Let $\psi_1, \ldots, \psi_p$ be the eigenvalues of G and let

$$-y_j^2/2 = \tilde{N}^{-1} \left\{ N_G \log (\psi_j/N_G) - (\psi_j - N_G) \right\}. \qquad (3.17)$$

We prove almost sure convergence by showing that [21, p. 151]

$$\text{Prob} \left\{ \cup_S \cup_{n \geq k} \left[\sum_{j=1}^{p} y_j^2/2 > \epsilon \right] \right\} \qquad (3.18)$$

$$\leq \sum_S \sum_{n \geq k} \text{Prob} \left[\sum_{j=1}^{p} y_j^2/2 > \epsilon \right] \to 0 \text{ as } k \to \infty.$$

Changing variables in the density of the eigenvalues of G [17], we obtain

$$\text{Prob}\left[\sum_{j=1}^{p} y_j^2/2 > \epsilon\right] = \prod_{j=1}^{p}\left[\Gamma(p-j+1)\,\Gamma(N_G-j+1)\right]^{-1}$$

$$\cdot N_G^{pN_G} e^{-pN_G} \int\cdots\int_{\Sigma y_j^2/2>\epsilon} \tilde{N}^p \exp\left\{-\tilde{N}\sum_{j=1}^{p} y_j^2/2\right\} \qquad (3.19)$$

$$\cdot \prod_{k=1}^{p-1}\prod_{j=k+1}^{p} (\psi_j^{-1} - \psi_k^{-1})^2 \prod_{j=1}^{p}\left(y_j/(\psi_j - N_G)\right) dy_1 \cdots dy_p \cdot$$

Since $\tilde{N}$, which appears in exponential in (3.19), goes to infinity at least as fast as the square root of the total number of data points, (3.18) can be shown to hold. Applying the same reasoning to the third term, we see that the right side of (3.13) becomes positive so that the probability of this subsequence occurring infinitely often is zero.

The rest of the proof consists of applying similar reasoning to the other cases. In each case we isolate a set S* for which the estimated spectral matrix does not belong to the neighborhood of the true spectral matrix and for which the number of points goes to infinity fast enough. For this set $\inf_F Q$ is shown to become positive and for the rest of the interval inf Q is shown to approach zero. If $\hat{S} \supset \tilde{S}^{(n)}$ and $\hat{F}_0 + uu^* \in \mathcal{N}_1$, then in the set $S^* = \left[f = f_B + 1\right]$ either $\hat{F}_0 \notin \mathcal{N}_0$ when $\hat{S} = \tilde{S}^{(n)}$ or $\hat{F}_0 + \hat{u}\hat{u}^* \notin \mathcal{N}_0$ when $\tilde{S}^{(n)}$ is a proper subset of $\hat{S}$. In either case, the previous reasoning applies. The case $\hat{S} \subset \tilde{S}^{(n)}$ can be approached in the same manner. In the other instance of frequency division, we consider first $\hat{S} \supset S_0^{(n)} - \tilde{S}^{(n)}$. If $\hat{F}_0 + \hat{u}\hat{u}^* \notin \mathcal{N}_0$, then the proof can be based on the set $S^* = S_0^{(n)} - \tilde{S}^{(n)}$. If $\hat{F}_0 + \hat{u}\hat{u}^* \in \mathcal{N}_0$, then $\hat{F}_0 \notin \mathcal{N}_1$ or $\hat{F}_0 + \hat{u}\hat{u}^* \notin \mathcal{N}_1$, so we can use the set $S^* = \left[f = f_B\right]$. The case $\hat{S} \subset S_0^{(n)} - \tilde{S}^{(n)}$ follows similarly. For time divisions, we treat the cases $\hat{S}$ contains more than half of $S_0^{(n)}$ separately from the cases $(S_0^{(n)} - \hat{S})$ contains more than half of $S_0^{(n)}$.

3.2 Distribution of the Criterion

Consistency is a property that informs us about the procedure when $u^*u > 0$. Two types of information can be obtained by studying the procedure when the population spectral matrix is uniform throughout the interval. First is the threshold to which the maximum dissimilarity should be compared in testing for stationarity and uniformity in frequency. Second is the bias associated with the procedure.

The joint distribution of the largest and smallest eigenvalues of $A_0^{-1}A$ has been derived [16, 17, 19]. From this result, the joint density of d(S) and $d(S_0 - S)$ for a given S can be obtained. The result is given in the following theorem.

Theorem 3: Let λ_1 and λ_p be the largest and smallest eigenvalues of $A_0^{-1}A$, and let $N_0 - p \geq N \geq p$. Under the null distribution ($F(f, t) = F_0$), we have

$$\text{Prob}\left\{u \leq \lambda_p \leq \lambda_1 \leq v\right\} = c_1 \det(b_{ij}), \tag{3.20}$$

where

$$b_{ij} = \int_u^v (x-\alpha)^{i+j-2} x^{N-p} (1-x)^{N_0 - N - p} dx, \tag{3.21}$$

$$c_1 = \prod_{i=1}^{p} \frac{\Gamma(N_0 - i + 1)}{\Gamma(N_0 - N - i + 1)\,\Gamma(N - i + 1)\,\Gamma(p - i + 1)}. \tag{3.22}$$

We choose $\alpha = 0$ or $\alpha = N/N_0$ for different purposes. The distribution of λ_1 for $N < p$ and the distribution of λ_p for $N > N_0 - p$ can be found similarly [17].

An asymptotic distribution for d(S) can be obtained by letting $\alpha = N/N_0$ in (3.21) and substituting asymptotic expansions for b_{ij} and the gamma functions.

Theorem 4: If N and $N_0 - N$ approach infinity in such a way that $N/N_0 \to \beta$ where $0 < \beta < 1$, we have under the null distribution

$$\text{Prob}\left\{0 \leq d(S),\ d(S_0 - S) \leq D\right\} \sim$$

$$\left(\prod_{i=1}^{p} \Gamma(i)\right)^{-1} \det\left((2\pi)^{-1/2} \int_{-D}^{D} y^{i+j-2} e^{-y^2/2}\, dy\right). \tag{3.23}$$

This distribution is shown in Figure 3 for $p = 1, 2, 4, 8, 16$ and 32. Since (3.23) does not depend on N or N_0, we see that in a limited sense our procedure is asymptotically unbiased under the null distribution.

Our test for uniformity of the population spectral matrix within a given interval rejects the hypothesis of uniformity if $\max_{S \in \mathcal{S}} d_e(S)$ is greater than some threshold. Since

$$\text{Prob}\left\{d(S) \text{ or } d(S_0 - S) > D\right\} \leq \text{Prob}\left\{\max_{S \in \mathcal{S}} d(S) > D\right\}$$

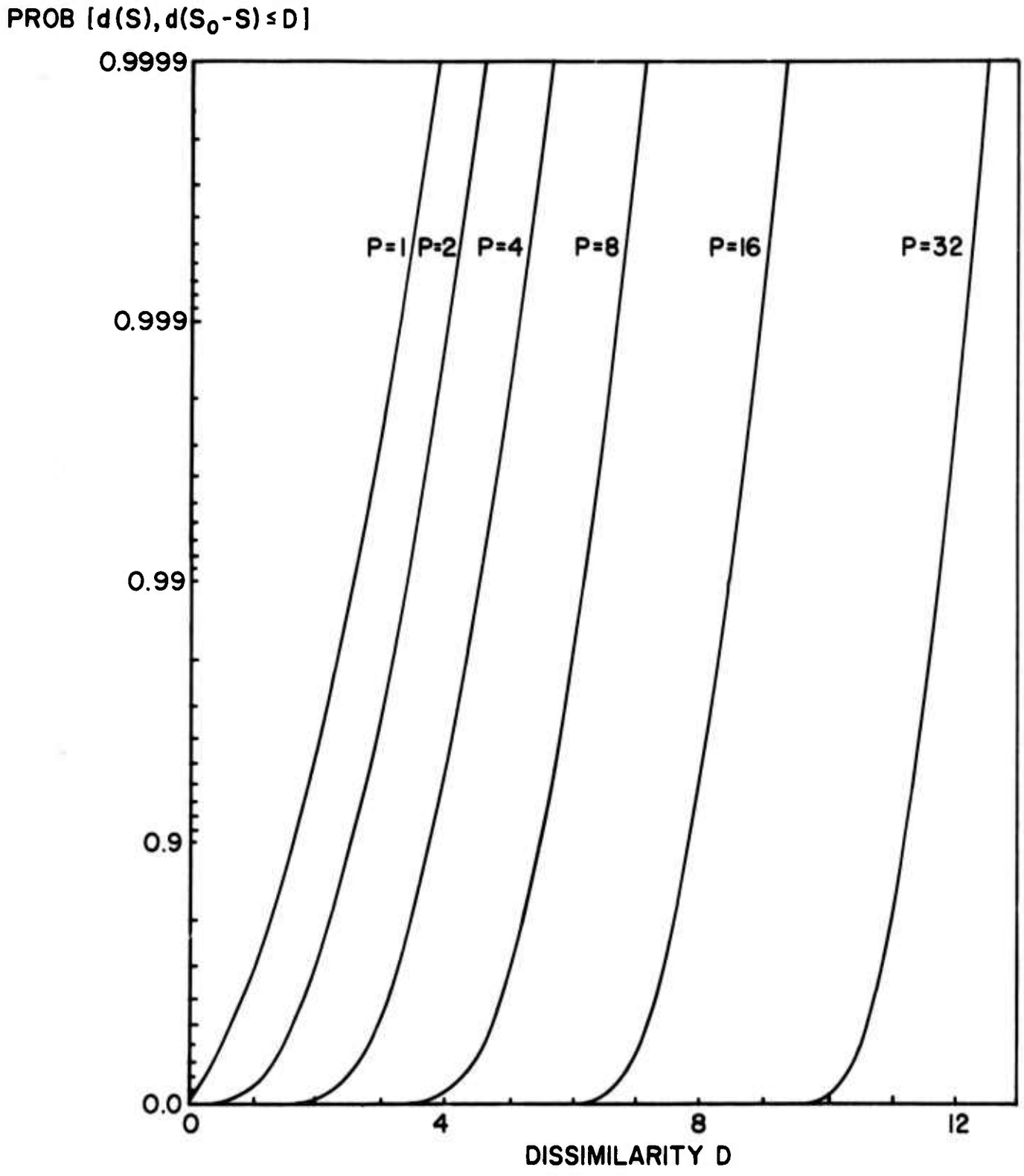
$$\leq m\ \text{Prob}\left\{d(S) \text{ or } d(S_0 - S) > D\right\}, \tag{3.24}$$

where m is the number of possible divisions of the interval, (3.23) gives approximate bounds on the level of significance for this test. As indicated below, N and N_0 must be large for the approximation in (3.23) to be close. Considering the values achieved by the maximum dissimilarity when the population spectral matrix is not uniform, we see that the upper bound given by (3.24) is not too high to be useful.

In order to obtain a better bound for the significance level of our test of uniformity as well as further insight into the procedure, we now derive an approximation to higher-order joint distributions of the dissimilarities. This result applies only to subintervals S_k and $S_0 - S_k$ for which $S_m \subset S_{m-1} \subset \ldots \subset S_1$. Also, the higher the order, the more effort is required in computing numerical values. However, joint distributions of any order can be used to obtain an upper bound on the significance level, with higher orders producing better bounds.

The following theorem gives the probability density under the null hypothesis of the eigenvalues of $A_0^{-1} A_k$, $k = 1, \ldots, m$, where

FIGURE 3. ASYMPTOTIC DISTRIBUTION FOR VARIOUS DIMENSIONS



$$A_k = \sum_{(f,\ t) \in S_k} a(f,\ t),\ k = 0,\ 1, \ldots, m \tag{3.25}$$

and $S_m \subset S_{m-1} \subset \ldots \subset S_1 \subset S_0$. From this, the joint distribution of $d(S_k)$, $d(S_0 - S_k)$, $k = 1, \ldots m$ is obtained in Theorem 6. The restriction $S_m \subset S_{m-1} \subset \ldots \subset S_1$ means that this theorem cannot be used to evaluate the dependence of d(S) for frequency divisions on d(S) for time divisions. The case of frequency and time divisions appears to be intractable. The result we derive applies only when all the roots of $A_0^{-1} A_k$ are greater than all the roots of $A_0^{-1} A_{k+1}$ for $k = 1, \ldots, m-1$. To approximate the density for all values of the roots, we observe that the probability that the sets of roots overlap approaches zero as the number of points in $S_k - S_{k+1}$ approaches infinity.

Theorem 5: Let $S_m \subset S_{m-1} \subset \ldots \subset S_1 \subset S_0$; let N_k be the number of points in S_k; and let N_m, $N_{m-1} - N_m$, $N_{m-2} - N_{m-1}, \ldots, N_2 - N_1$, $N_0 - N_1 \geq p$. For $k = 1, \ldots, m$, let the eigenvalues of $A_0^{-1} A_k$ in descending order be $\lambda_{k1}, \lambda_{k2}, \ldots, \lambda_{kp}$. Let the a(f, t) be distributed as above with the same population spectral matrix throughout S_0. Then, when $\lambda_{kp} \geq \lambda_{(k+1)1}$, for $k = 1, \ldots, m-1$, the joint density of the eigenvalues is

$$\begin{aligned} p(\lambda_{11}, \ldots, \lambda_{mp}) = c_m \prod_{i=1}^{p} \left\{ \lambda_{mi}^{N_m - p} (1 - \lambda_{1i})^{N_0 - N_1 - p} \right\} \\ \cdot \prod_{k=1}^{m-1} \left\{ \det \left[\left(\lambda_{ki} - \lambda_{(k+1)j} \right)^{N_k - N_{k+1} - 1} \right] \right\} \\ \cdot \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} (\lambda_{1i} - \lambda_{1j}) (\lambda_{mi} - \lambda_{mj}), \end{aligned} \tag{3.26}$$

where

$$\begin{aligned} c_m = \prod_{i=1}^{p} \left\{ \frac{\Gamma(N_0 - i + 1)}{\Gamma(N_0 - N_1 - i + 1)\, \Gamma(N_m - i + 1)\, \Gamma(p - i + 1)} \right. \\ \left. \cdot \prod_{k=1}^{m-1} \frac{1}{\Gamma(N_k - N_{k+1})} \right\}. \end{aligned} \tag{3.27}$$

Proof: Using [17], we obtain the joint distribution of

$$B_k = A_0^{-1/2} A_k A_0^{-1/2}, \quad k=1,\ldots,m \tag{3.28}$$

which is

$$p(B_1, B_2, \ldots, B_m)$$

$$= \frac{\tilde{\Gamma}_p(N_0)\left(\det B_m\right)^{N_m - p}\left(\det (I - B_1)\right)^{N_0 - N_1 - p}}{\tilde{\Gamma}_p(N_m)\,\tilde{\Gamma}_p(N_0 - N_1)} \tag{3.29}$$

$$\cdot \prod_{k=1}^{m-1} \left\{ \frac{\left(\det (B_k - B_{k+1})\right)^{N_k - N_{k+1} - p}}{\tilde{\Gamma}_p(N_k - N_{k+1})} \right\},$$

where $\tilde{\Gamma}_p(N)$ is the complex multivariate Gamma function,

$$\tilde{\Gamma}_p(N) = \pi^{p(p-1)/2} \prod_{i=1}^{p} \Gamma(N - i + 1). \tag{3.30}$$

Using [17], we make the following sequence of transformations. For $k=1, 2, \ldots, m$, let $\Lambda_k = \text{diag}\,(\lambda_{k1}, \lambda_{k2}, \ldots, \lambda_{kp})$ and let U_k be unitary matrices. First, we transform B_1 to $U_1 \Lambda_1 U_1^*$ and then transform $U_1^* B_k U_1$ to $B_k^{(1)}$ for $k=2,\ldots,m$. Second, we transform $B_2^{(1)}$ to $U_2 \Lambda_2 U_2^*$ and then transform $U_2^* B_k^{(1)} U_2$ to $B_k^{(2)}$ for $k=3,\ldots,m$. We continue this until we transform $B_m^{(m-1)}$ to $U_m \Lambda_m U_m^*$. Integrating over U_k for the case $\lambda_{kp} \geq \lambda_{(k+1)1}$, we obtain

$$p(\lambda_{11}, \ldots, \lambda_{mp}) = c_m \left(\det \Lambda_m\right)^{N_m - p} \left(\det (I - \Lambda_1)\right)^{N_0 - N_1 - p} \tag{3.31}$$

(cont)

$$\cdot \prod_{k=1}^{m} \left\{ \frac{\int_{U(p)} \left(\det (\Lambda_k - U \Lambda_{k+1} U^*)\right)^{N_k - N_{k+1} - p} (dU)}{\prod_{i=1}^{p} \left[\Gamma(N_k - N_{k+1} - i + 1)\,\Gamma(p - i + 1) / \Gamma(N_k - N_{k+1})\right]} \right\}$$

$$\cdot \prod_{k=1}^{m} \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} (\lambda_{ki} - \lambda_{kj})^2, \qquad (3.31)\ (\text{cont})$$

where (dU) is the invariant measure on the unitary group U(p) normalized to make the total measure unity [14]. Note that when $\lambda_{kp} < \lambda_{(k+1)1}$ the range of integration of U is limited by the requirement that $\Lambda_k - U\Lambda_{k+1}U^*$ be positive semi-definite.

Using [14], we can evaluate the integrals over U. We have

$$\int_{U(p)} \left[\det\ (I - AUBU^*)\right]^{n-p} (dU) = {}_1\tilde{F}_0\,(-n+p;\ A,\ B), \qquad (3.32)$$

where $A = \text{diag}\ (\alpha_1, \ldots, \alpha_p)$, $B = \text{diag}\ (\beta_1, \ldots, \beta_p)$, and ${}_1\tilde{F}_0$ is a hypergeometric function of matrix argument defined in [14]. Using the definitions provided in [14], we obtain

$${}_1\tilde{F}_0\,(-n+p;\ A,\ B) =$$

$$\sum_{k=0}^{\infty} \sum_{k_1 \geq \ldots \geq k_p} \prod_{i=1}^{p} \left[\frac{\Gamma(p-i+1)\ (-n+p-i+1)_{k_i}}{\Gamma(k_i+p-i+1)}\right] \qquad (3.33)$$

$$\cdot \frac{\det\left(\alpha_i^{\,k_j+p-j}\right) \det\left(\beta_i^{\,k_j+p-j}\right)}{\det\left(\alpha_i^{\,p-j}\right) \det\left(\beta_i^{\,p-j}\right)},$$

where the sum over $k_1 \geq \ldots \geq k_p$ is the sum over the partitions of k and $(n)_j = (n) \cdot (n+1) \ldots (n+j-1)$. To sum (3.33), we need the identity

$$\det\left(\alpha_i^{\,k_j+p-j}\right) \det\left(\beta_i^{\,k_j+p-j}\right)$$

$$= \sum_{r_1, \ldots, r_p} \text{sign}\ (r_1, \ldots, r_p) \sum_{s_1, \ldots, s_p} \prod_{i=1}^{p} \left(\alpha_{r(s_i)}\, \beta_{s_i}\right)^{k_i+p-i}, \qquad (3.34)$$

where the summations are over the permutations of (1, 2,...,p), $r(s_i)$ denotes the result of successive permutations of i by r and s, and sign $(r_1,\ldots,r_p)$ is +1 (-1) if the permutation is even (odd). By summing the binomial series, we obtain

$$\det\left(\left(\alpha_i^{\,p-j}\right)\det\left(\beta_i^{\,p-j}\right)\right)\ {}_1\tilde{F}_0\,(-n+p;\,A,\,B)$$

$$= \left[\prod_{i=1}^{p}\Gamma(p-i+1)\right]\sum_{r_1,\ldots,r_p}\operatorname{sign}(r_1,\ldots,r_p)$$

$$\cdot\sum_{k=0}^{\infty}\sum_{k_1\geq\ldots\geq k_p}\sum_{s_1,\ldots,s_p}\prod_{i=1}^{p}\left[\frac{(-n+p-i+1)_{k_i}}{\Gamma(k_i+p-i+1)}\left(\alpha_{r(s_i)}\beta_{s_i}\right)^{k_i+p-i}\right] \tag{3.35}$$

$$= \prod_{i=1}^{p}\left[\frac{\Gamma(p-i+1)}{(-n+1)_{p-i}}\right]\sum_{r_1,\ldots,r_p}\operatorname{sign}(r_1,\ldots,r_p)\prod_{i=1}^{p}\left(1-\alpha_{r_i}\beta_i\right)^{n-1}$$

$$= (-1)^{p(p-1)/2}\prod_{i=1}^{p}\left[\Gamma(p-i+1)\,\Gamma(n-i+1)/\Gamma(n)\right]\det\left(\left(1-\alpha_j\beta_i\right)^{n-1}\right).$$

Substituting (3.32) and (3.35) into (3.31) completes the proof.

Theorem 5 provides a formula for the density that is exact when the roots from different intervals do not overlap. We now integrate this density.

Theorem 6: If the conditions of Theorem 5 hold and if $v_1 \geq u_1 \geq v_2 \geq u_2 \ldots \geq v_m \geq u_m$, then

$$\text{Prob}\left\{v_k \geq \lambda_{k1},\ldots,\lambda_{kp} > u_k;\ k=1,\ldots,m\right\} = c_m \det(a_{ij}) \tag{3.36}$$

where

$$a_{ij} = \int_{u_1}^{v_1} \int_{u_2}^{v_2} \cdots \int_{u_m}^{v_m} (\lambda_m - \alpha_1)^{p-i} (\lambda_1 - \alpha_2)^{p-j} \lambda_m^{N_m - p} \tag{3.37}$$

$$\cdot \left(1 - \lambda_1\right)^{N_0 - N_1 - p} \prod_{k=1}^{m-1} \left(\lambda_k - \lambda_{k+1}\right)^{N_k - N_{k+1} - 1} d\lambda_m \cdots d\lambda_1 .$$

The constants α_1 and α_2 can be chosen at our convenience.

Proof: Since, for any i, j, and k, λ_{ki} and λ_{kj} can be interchanged without changing the value of the expression for $p(\lambda_{11}, \ldots, \lambda_{mp})$ given in (3.26), we have

$$\int_{u_1}^{v_1} \int_{u_1}^{\lambda_{11}} \cdots \int_{u_1}^{\lambda_{m(p-1)}} p(\lambda_{11}, \ldots, \lambda_{mp})\, d\lambda_{mp} \cdots d\lambda_{11}$$

$$= \left[\Gamma(p+1)\right]^{-m} \int_{u_1}^{v_1} \int_{u_1}^{v_1} \cdots \int_{u_m}^{v_m} p(\lambda_{11}, \ldots, \lambda_{mp})\, d\lambda_{mp} \cdots d\lambda_{11} . \tag{3.38}$$

Using the identity

$$\prod_{i=1}^{p-1} \prod_{j=i+1}^{p} (\lambda_{mi} - \lambda_{mj}) = \det \left(\left(\lambda_{mi} - \alpha_1\right)^{p-j} \right), \tag{3.39}$$

and expanding the determinants, we have

$$\prod_{i=1}^{p} \lambda_{mi}^{N_m - p} \det \left(\left(\lambda_{(m-1)i} - \lambda_{mj}\right)^{N_{m-1} - N_m - 1} \right)$$

$$\cdot \prod_{i=1}^{p-1} \prod_{j=i+1}^{p} (\lambda_{mi} - \lambda_{mj}) \tag{3.40}$$

$$= \sum_{r_1, \ldots, r_p} \sum_{s_1, \ldots, s_p} \operatorname{sign}(s_1, \ldots, s_p)$$

$$\cdot \prod_{i=1}^{p} \lambda_{mr_i}^{N_m - p} \left(\lambda_{(m-1)i} - \lambda_{mr_i}\right)^{N_{m-1} - N_m - 1} (\lambda_{mr_i} - \alpha_1)^{p - s_i} ,$$

where $r_1,\ldots,r_p$ and $s_1,\ldots,s_p$ are permutations of $1,\ldots,p$. Integrating (3.40) over $u_m \leq \lambda_{mi} \leq v_m$, $i=1,\ldots,p$, we obtain

$$\Gamma(p+1)\sum_{s_1,\ldots,s_p} \operatorname{sign}(s_1,\ldots,s_p)\prod_{i=1}^{p}\left[\int_{u_m}^{v_m} \lambda_m^{N_m-p}\right.$$

$$\left.\cdot\left(\lambda_{(m-1)i}-\lambda_m\right)^{N_{m-1}-N_m-1}\left(\lambda_m-\alpha_1\right)^{p-s_i}\,d\lambda_m\right] \tag{3.41}$$

$$=\Gamma(p+1)\det\left(\int_{u_m}^{v_m} \lambda_m^{N_m-p}\left(\lambda_{(m-1)i}-\lambda_m\right)^{N_{m-1}-N_m-1}\right.$$

$$\left.\cdot\left(\lambda_m-\alpha_1\right)^{p-j}\,d\lambda_m\right).$$

Repeating this process proves the theorem.

An approximation for the distribution of λ_{k1}, $k=1,\ldots,m$ valid for large N_0-N_1, $N_1-N_2,\ldots,N_{m-1}-N_m$, N_m can be obtained by extending the lower limits in (3.37) into the region where the density is not valid. When the contribution from this region is small, this approximation is useful. We obtain

$$\text{Prob}\left\{v_k \geq \lambda_{k1};\ k=1,\ldots,m\right\} \sim c_m \det\left(a'_{ij}\right), \tag{3.42}$$

where

$$a'_{ij}=\int_0^{v_m}\int_{\lambda_m}^{v_{m-1}}\cdots\int_{\lambda_2}^{v_1}\left(\lambda_m-\alpha_1\right)^{p-i}\left(\lambda_1-\alpha_2\right)^{p-j} \tag{3.43}$$

$$\cdot\,\lambda_m^{N_m-p}\left(1-\lambda_1\right)^{N_0-N_1-p}\prod_{k=1}^{m-1}\left(\lambda_k-\lambda_{k+1}\right)^{N_k-N_{k+1}-1}$$

$$\cdot\,d\lambda_1\ldots d\lambda_m.$$

It can be shown that (3.42) gives the correct total probability, $\text{Prob}\left\{1 \geq \lambda_{k1};\ k=1,\ldots,m\right\}=1$.

For the case $p=1$, (3.36) and (3.37) can be recognized as the probability that the sample quantiles of order N_1/N_0, $N_2/N_0, \ldots, N_m/N_0$ lie within the given limits for independent samples from a uniform (0, 1) distribution. Thus, the relation of our procedure to the test presented in [8] for a single time series can be seen. This test is based on the maximum of $\left|\lambda_{k1} - k/N_0\right|$ over $1 \leq k < N_0$. Thus, this test chooses $S_k = \left\{(f, t) \,\middle|\, f_L \leq f \leq f_L + k - 1,\ t = t_L\right\}$ and uses a different function of λ_{k1} as the measure of dissimilarity.

Setting $\alpha_1 = N_m/N_0$ and $\alpha_2 = N_1/N_0$ in (3.37), we can obtain an asymptotic expansion for a_{ij} and thus an asymptotic distribution for $d(S_k)$, $d(S_0 - S_k)$, $k = 1, \ldots, m$. The expansion of a_{ij} is nearly the same as that obtained in the asymptotic distribution of sample quantiles [6, p. 201].

In order to obtain moments, we consider an equivalent formulation of our procedure. Let $\lambda'_{k1}, \lambda'_{k2}, \ldots, \lambda'_{kp}$ be the eigenvalues of $A_0^{-1} A_k$ taken in a random order and let

$$y_{kj} = \frac{\left(\lambda'_{kj} - N_k/N_0\right)}{\left|\lambda'_{kj} - N_k/N_0\right|} \left\{ 2 N_k \log\left[\frac{N_k}{N_0 \lambda'_{kj}}\right] + 2 (N_0 - N_k) \log\left[\frac{N_0 - N_k}{N_0(1 - \lambda'_{kj})}\right] \right\}^{1/2} . \tag{3.44}$$

Our procedure is equivalent to selecting $S_k \in S$ if $\left|y_{kj}\right| = \max_{k,j} \left|y_{kj}\right|$ for some value of j. We obtain expressions for the second moments of y_{kj}.

Theorem 7: Let y_{1j}, y_{2j} be defined as in (3.44). If N_0, N_1, N_2, approach infinity in such a way that $N_1/N_0 \longrightarrow \beta_1$ and $N_2/N_0 \longrightarrow \beta_2$, $0 < \beta_2 < \beta_1 < 1$, then we have

$$Ey_{1j},\ Ey_{2j} \sim 0$$

$$Ey_{1j}^2,\ Ey_{2j}^2 \sim p$$

$$Ey_{1i}y_{1j},\ Ey_{2i}y_{2j} \sim -1 \qquad i \neq j \tag{3.45}$$

$$Ey_{1i}y_{2j} \sim p^{-1}\left[\beta_2(1-\beta_1)/(\beta_1(1-\beta_2))\right]^{1/2}.$$

Proof: We first derive the moments of λ'_{kj} and then use the expansion

$$y_{kj} \sim \left[N_0/(\beta_k(1-\beta_k))\right]^{1/2}(\lambda'_{kj}-\beta_k). \tag{3.46}$$

Proceeding as we did in the proof of Theorem 6, we obtain

$$p(\lambda'_{11},\ldots,\lambda'_{2p}) =$$

$$c_2\left[\Gamma(p+1)\right]^{-2}(-1)^{p(p-1)/2}\sum_{q_1,\ldots,q_p}\sum_{r_1,\ldots,r_p}\sum_{s_1,\ldots,s_p}$$

$$\text{sign}\,(s_1,\ldots,s_p)\ \prod_{i=1}^{p}\left\{\left(\lambda'_{2q_i}\right)^{N_1-i}\left(\lambda'_{1r_i}-\lambda'_{2q_i}\right)^{N_1-N_2-1}\right. \tag{3.47}$$

$$\left.\cdot\left(1-\lambda'_{1r_i}\right)^{N_0-N_1-s_i}\right\}.$$

We integrate applying the approximation used in (3.42) and (3.43) to obtain

$$E\lambda'_{2n} \sim (-1)^{p(p-1)/2}\,c_2\left[\Gamma(p+1)\right]^{-1}$$

$$\cdot\sum_{q_1,\ldots,q_p}\det\left[\frac{\Gamma(N_2+\delta(n,q_i)-i+1)\,\Gamma(N_1-N_2)\,\Gamma(N_0-N_1-j+1)}{\Gamma(N_0+\delta(n,q_i)-i-j+2)}\right] \tag{3.48}$$

$$=(-1)^{p(p-1)/2}\prod_{i=1}^{p}\left|\Gamma(p-i+1)\right|^{-1}\frac{1}{p}\sum_{k=1}^{p}(N_2-k+1)\cdot\det\left[\frac{\Gamma(N_0-j+1)}{\Gamma(N_0+\delta(i,k)-i-j+2)}\right],$$

where $\delta(i, j)$ is the Kronecker delta. Making use of the identity

$$N_0 \det\left[\Gamma(N_0 - i + 1)/\Gamma(N_0 + \delta(i, 1) - i - j + 2)\right] =$$

$$= \det\left[\left(N_0 + \delta(i, 1) - i\right)^{j-1}\right] = (-1)^{p(p-1)/2} \, p \prod_{i=1}^{p} \Gamma(p - i + 1), \tag{3.49}$$

we see that $E\lambda'_{2j} \sim N_2/N_0$. The derivation of $E(1 - \lambda'_{1j})$, $E(\lambda'_{2i} \lambda'_{2j})$, and $E((1 - \lambda'_{1i})(1 - \lambda'_{ij}))$ proceeds similarly.

Consider finally the derivation of $E(\lambda'_{2i}(1 - \lambda'_{1j}))$, where without loss of generality we let $i = j = 1$. Integrating and replacing r_i by $r(s_i)$, we obtain

$$E\left(\lambda'_{21}(1 - \lambda'_{11})\right) = (-1)^{p(p-1)/2} \, c_2 \left[\Gamma(p+1)\right]^{-2}$$

$$\cdot \sum_{q_1, \ldots, q_p} \sum_{r_1, \ldots, r_p} \sum_{s_1, \ldots, s_p} \operatorname{sign}(s_1, \ldots, s_p)$$

$$\cdot \prod_{i=1}^{p} \left\{ \frac{\Gamma(N_2 + \delta(1, q_i) - i + 1) \, \Gamma(N_1 - N_2) \, \Gamma(N_0 - N_1 + \delta(1, r(s_i)) - s_i + 1)}{\Gamma(N_0 + \delta(1, q_i) + \delta(1, r(s_i)) - i - s_i + 2)} \right\} \tag{3.50}$$

$$= (-1)^{p(p-1)/2} \prod_{i=1}^{p} \left[\Gamma(p - i + 1)\right]^{-1} \frac{1}{p^2} \sum_{k=1}^{p} \sum_{k'=1}^{p}$$

$$(N_2 - k + 1)(N_0 - N_1 - k' + 1) \det\left\{ \frac{\Gamma(N_0 - i + 1)}{\Gamma(N_0 + \delta(i, k) + \delta(j, k') - i - j + 2)} \right\}$$

In (3.50), the determinant is non-zero only when $k = k' = 1$. The determinant in (3.50) equals N_0^{-1} times the following determinant which we evaluate by expanding about the elements of the first column and evaluating the minors using results similar to (3.49). We obtain

$$\det\left\{\frac{\Gamma(N_0+\delta(i,1)-i+1)}{\Gamma(N_0+\delta(i,1)+\delta(j,1)-i-j+2)}\right\} = \prod_{i=1}^{p}(N_0+\delta(i,1)-i)$$

$$\cdot\sum_{m=1}^{p}\left[\frac{(-1)^{m-1}}{(N_0+\delta(m,1)-m+1)(N_0+\delta(m,1)-m)}\right]$$

$$\cdot\prod_{\substack{i=1\\ i\neq m}}^{p-1}\prod_{\substack{j=i+1\\ j\neq m}}^{p}(-j-\delta(i,1)+i)$$

$$=(-1)^{p(p-1)/2}\prod_{i=1}^{p}\left[\Gamma(p-i+1)(N_0-\delta(i,1)-i)\right] \tag{3.51}$$

$$\cdot\left\{\frac{p}{\Gamma(p)}\sum_{m=1}^{p}(-1)^{p-m}\binom{p-1}{m-1}\left[(N_0-m+1)(N_0-m)\right]^{-1}\right.$$

$$\left.-\frac{p}{\Gamma(p+1)}\sum_{m=0}^{p}(-1)^{p-m}\binom{p}{m}\left[(N_0-m+1)(N_0-m)\right]^{-1}\right\}.$$

The summations on the right side of (3.51) are evaluated using the identity

$$\sum_{m=1}^{p}(-1)^{p-m}\binom{p-1}{m-1}\left[(N_0-m+1)(N_0-m)\right]^{-1}$$

$$=\int_0^1\int_0^y x^{N_0-p-1}(1-x)^{p-1}\,dxdy \tag{3.52}$$

$$=\left[\Gamma(p)\,\Gamma(N_0-p+1)/\Gamma(N_0)\right]\sum_{m=1}^{p}\left[(N_0-p+m)(N_0-p+m-1)\right]^{-1}$$

and the equivalent identity obtained by replacing p by p+1. Equation (3.45) follows from the O(1) and $O(N_0^{-1})$ terms of the result.

3.3 Small - Sample Bias

We would like to have a procedure that under the null hypothesis choses any division with equal probability. Alternatively, a procedure that tends to divide the interval into equal-sized subintervals is preferable to other tendencies. Neither of

these properties hold for our procedure, and we have excluded certain values of d(S) from our selection to improve its behavior. As shown by the form of the moments in (3.45), our procedure does not select divisions with equal probability even asymptotically. There is also a small-sample bias. We investigate this problem by studying the univariate and bivariate distributions derived above and by simulation.

Under the null hypothesis, the maximum of d(S) tends to occur for large N. In the extreme, we see that $d(S) = \infty$ for $N > N_0 - p$. One indication of this bias is given in the following:

Theorem 8: When $N > N_0/2$, we have under the null hypothesis

$$\text{Prob}\left\{d(S) \geq d(S_0 - S) \geq \alpha\right\} > \text{Prob}\left\{d(S_0 - S) \geq d(S) \geq \alpha\right\}, \tag{3.53}$$

where α is any constant.

Proof: Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of $A_0^{-1} A$ in decreasing order. Dropping the unneeded subscript, let y_j be defined as in (3.44). Since this function is monotonic, we can consider λ_j to be a function of y_j, $\lambda_j = \lambda(y_j)$. Let $\lambda_j^{(1)} = \lambda^{(1)}(y_j) = d\lambda_j/dy_j$. Using [17], we obtain

$$\begin{aligned} &\text{Prob}\left\{d(S) \geq d(S_0 - S) \geq \alpha\right\} = \text{Prob}\left\{y_1 \geq -y_p \geq \alpha\right\} \\ &= \int_\alpha^\infty p_N(y_1)\, dy_1, \end{aligned} \tag{3.54}$$

where

$$\begin{aligned} p_N(y_1) = \frac{c_1}{\Gamma(p)} &\left[\left(\frac{N}{N_0}\right)^N \left(\frac{N_0 - N}{N}\right)^{N_0 - N}\right]^p e^{-y_1^2/2} \quad \frac{\lambda_1^{(1)}}{\lambda_1(1-\lambda_1)} \\ &\cdot \int_{-y_1}^{y_1} \cdots \int_{-y_1}^{y_1} \exp\left\{-\frac{1}{2}\sum_{j=2}^{p} y_j^2\right\} \end{aligned} \tag{3.55 (cont)}$$

$$\cdot \prod_{j=2}^{p} \left(\frac{1}{\lambda_j} - \frac{1}{\lambda_1}\right)\left(\frac{1}{1-\lambda_1} - \frac{1}{1-\lambda_j}\right) \prod_{k=2}^{p-1} \prod_{j=k+1}^{p} (\lambda_k - \lambda_j)^2 \tag{3.55}$$

$$\cdot \prod_{j=2}^{p} \lambda_j^{(1)} \quad dy_2 \cdots dy_p . \tag{cont}$$

If in (3.44), we exchange N with $N_0 - N$, we obtain another function of y_j which we denote $\psi_j = \psi(y_j)$. Since

$$\lambda(y_j) = 1 - \psi(-y_j); \quad \lambda^{(1)}(y_j) = \psi^{(1)}(-y_j) \tag{3.56}$$

we can obtain an expression for $p_{N_0 - N}(y_1)$ by replacing λ_1 with ψ_1 and $\lambda_1^{(1)}$ with $\psi_1^{(1)}$ in (3.55). We complete the proof by showing that $p_N(y_1) > p_{N_0 - N}(y_1)$. The curve $(\lambda(y_1), \psi(y_1))$ passes through $(N/N_0, (N_0 - N)/N_0$ and (1, 1), and at $y_1 = 0$, $\lambda_1^{(1)} = \psi_1^{(1)}$. Since

$$\frac{\psi_1^{(1)}}{\lambda_1^{(1)}} = \left[\frac{N_0\lambda_1 - N}{N_0\psi_1 - (N_0 - N)}\right]\left[\frac{\psi_1(1-\psi_1)}{\lambda_1(1-\lambda_1)}\right], \tag{3.57}$$

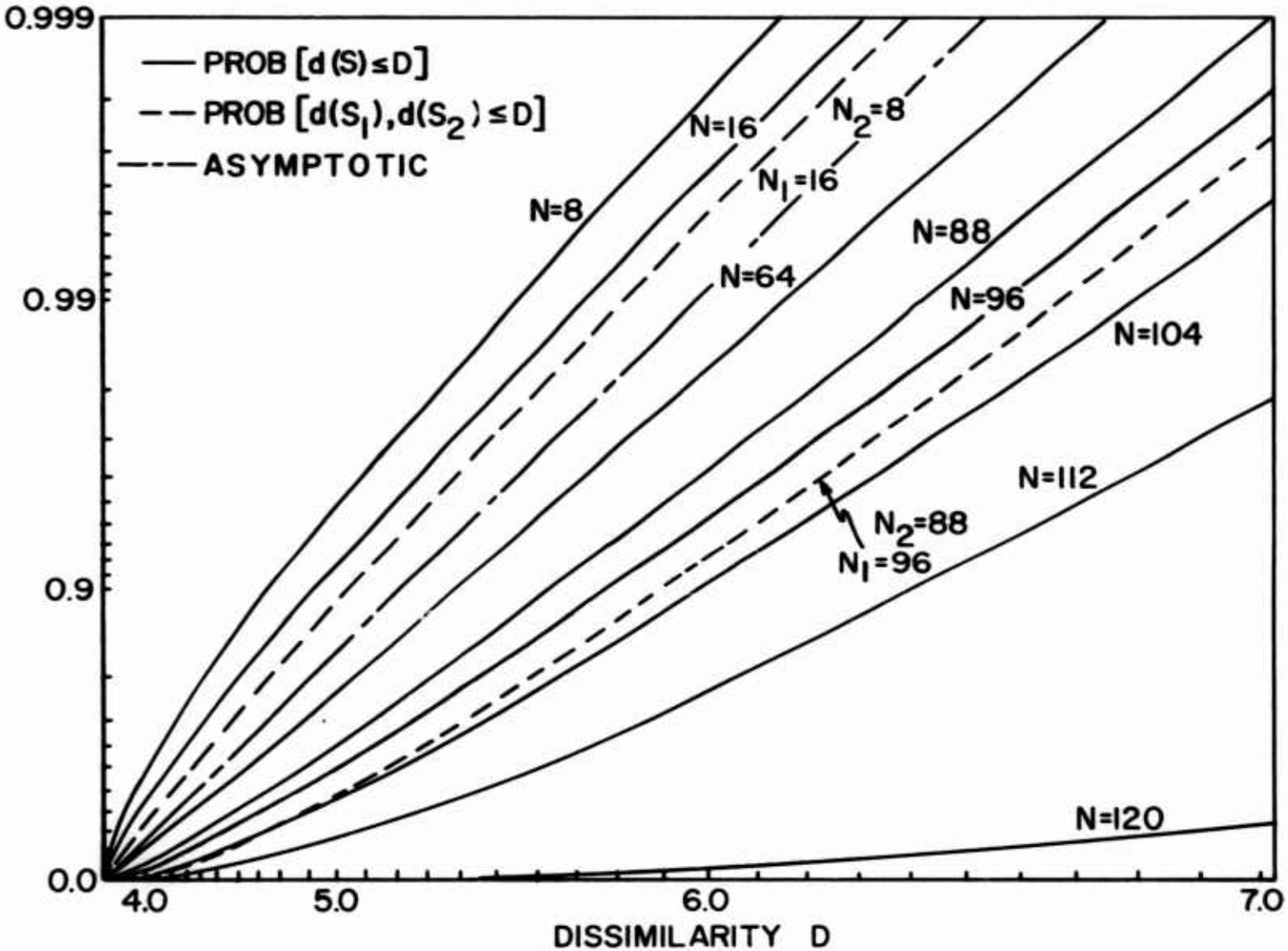
we see that $\psi_1^{(1)}/\lambda_1^{(1)} > 1$ if $N_0\lambda_1 - N > N_0\psi_1 - (N_0 - N)$ and $\psi_1^{(1)}/\lambda_1^{(1)} < 1$ if $\psi_1 > \lambda_1$. Thus, for $y_1 \geq 0$,

$$\lambda_1 \geq \psi_1; \quad N_0\lambda_1 - N \leq N_0\psi_1 - (N_0 - N). \tag{3.58}$$

This completes the proof.

Figure 4 gives Prob $\{d(S) \leq D\}$ and Prob $\{d(S_1)$ and $d(S_2) \leq D\}$ under the null hypothesis for $N_0 = 128$ and various N, N_1, and N_2. The horizontal scale is a non-linear function of D chosen so that asymptotic distribution of d(S) plots as the identity function. We make three observations. First, the probability of d(S) exceeding a given value increases with N as does the probability of $d(S_1)$ or $d(S_2)$ exceeding a given value. This fact agrees with the simulation result that the largest d(S) tends to occur for intervals with large N. Second, for large D, the sum of the probability of $d(S_1) > D$ and the probability of $d(S_2) > D$ is not much larger

FIGURE 4. UNIVARIATE AND BIVARIATE DISTRIBUTIONS FOR EIGHT SERIES



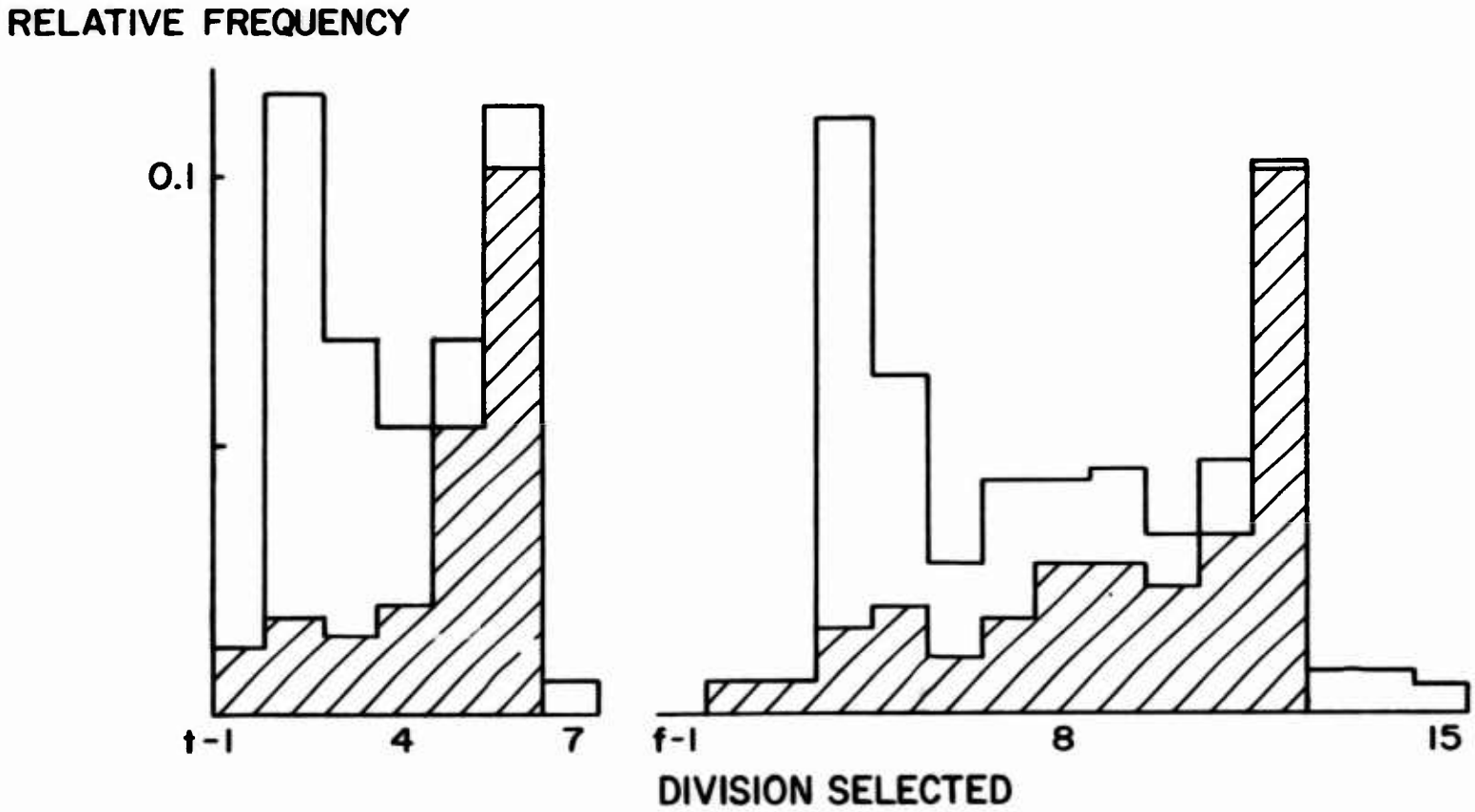
than the probability of $d(S_1)$ or $d(S_2) > D$. This indicates that the univariate distributions can be used to obtain an adequately tight upper bound on the level for testing uniformity. Finally, for a given D the value of the distribution for $N = 64$ is smaller than the value of the asymptotic distribution. This fact and the results for the other values of N indicate that the asymptotic distribution does not give true upper bounds on the level of significance. This also indicates that our procedure might tend to divide small intervals in preference to large intervals.

To obtain results on the small-sample bias that include all effects, we performed 500 replications of one step of our procedure. The data were identically-distributed, 8-dimensional, complex-Gaussian vectors for the interval $S_0 = \{(f, t) \mid 1 \leq f \leq 16, 1 \leq t \leq 8\}$. Figure 5 shows the histogram of divisions chosen. For the time division t - j, the shaded contribution to the relative frequency resulted from $d_e(S)$ being largest for $S = \{1 \leq f \leq 16, 1 \leq t \leq j\}$, and the unshaded contribution resulted from $d_e(S)$ being largest for $S = \{1 \leq f \leq 16, j < t \leq 8\}$. The results for the frequency divisions are displayed similarly. The tendency for the maximum of $d_e(S)$ to correspond to an interval with large N is clearly seen.

Our procedure excludes d(S) when $N > N_0 - 4p$. This is why divisions t-7, f-14, f-15, and f-16 have no shaded contribution and why division t-1, f-1, f-2, and f-3 have no unshaded contributions. Without this exclusion, the simulation results were much more biased. The two intervals with $N = 120$ were chosen 473 times out of 500.

The simulation results can be used to illustrate an upper bound on the level of significance for our test of uniformity. In the simulation, 9% of the 500 values of $\max_{S \in \mathcal{S}} d_e(S)$ exceeded 6.5. Using (3.20), we obtain 0.12 for the sum of the probabilities of $[d(S) > 6.5]$ for the 36 intervals examined in the simulation. Thus, this upper bound is apparently tight enough for most purposes. If we multiply the asymptotic probability of $[d(S) > 6.5]$ by 22, the total number of divisions, we obtain 0.03. Thus, in this case, the asymptotic distribution is far enough above the true distribution for most N that it underestimates the significance level.

FIGURE 5. SIMULATION FOR NULL DISTRIBUTION



The usefulness of an upper bound based on (3.20) can be expected to vary with p. In particular, since (3.45) seems to indicate that the dependence of $d(S_1)$ on $d(S_2)$ decreases with increasing p, such bounds may not be as tight for smaller p.

4. SIMULATION

Beside the analytical complexity, the broad range of alternatives to the null hypothesis makes precise specification of the performance of our procedure in all situations nearly impossible. In this section, we present the results of a simulation for a simple situation for which the correct behavior is easily seen. A simulation of a realistic sonar situation has been presented [20]. The simulation presented in this section shows that the sensitivity of our procedure increases with p, the number of time series analyzed. This suggests that determination of smoothing for each single time series in a multiple time series followed by application of the result to the analysis of the multiple time series is not as sensitive as our procedure. This simulation also provides information on the magnitude of the differences our procedure will detect.

The data we created were 16-component, complex-Gaussian vectors for the interval $S_0 = \{(f, t) \mid 1 \leq f \leq 40,\ 1 \leq t \leq 16\}$. The population spectral matrix was

$$F = \begin{cases} u_1 u_1^* + u_2 u_2^* + u_3 u_3^* + u_4 u_4^* + I & \text{for } 1 \leq f \leq 8 \\ u_2 u_2^* + u_3 u_3^* + u_4 u_4^* + I & \text{for } 9 \leq f \leq 16 \\ u_3 u_3^* + u_4 u_4^* + I & \text{for } 17 \leq f \leq 24 \\ u_4 u_4^* + I & \text{for } 25 \leq f \leq 32 \\ I & \text{for } 33 \leq f \leq 40, \end{cases} \tag{4.1}$$

where

$$u_1 = \frac{1}{2^{1/2}} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}, \quad u_2 = \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}, \quad u_3 = \frac{1}{8^{1/2}} \begin{pmatrix} 1 \\ i \\ -1 \\ -i \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}, \quad u_4 = \frac{1}{4} \begin{pmatrix} 1 \\ -i \\ -1 \\ i \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}. \tag{4.2}$$

Because u_1, u_2, u_3, and u_4 are mutually orthogonal, the population eigenvalues are easily determined.

We created 20 independent sets of data. We applied the procedure to each single time series, to consecutive pairs of time series, etc. to obtain 320 independent relications for p = 1, 160 independent replications for p = 2, 80 for p = 4, 40 for p = 8, and 20 for p = 16. The results for different values of p are dependent, a fact which does not make this simulation unsuitable for determining the increase in the sensitivity with p.

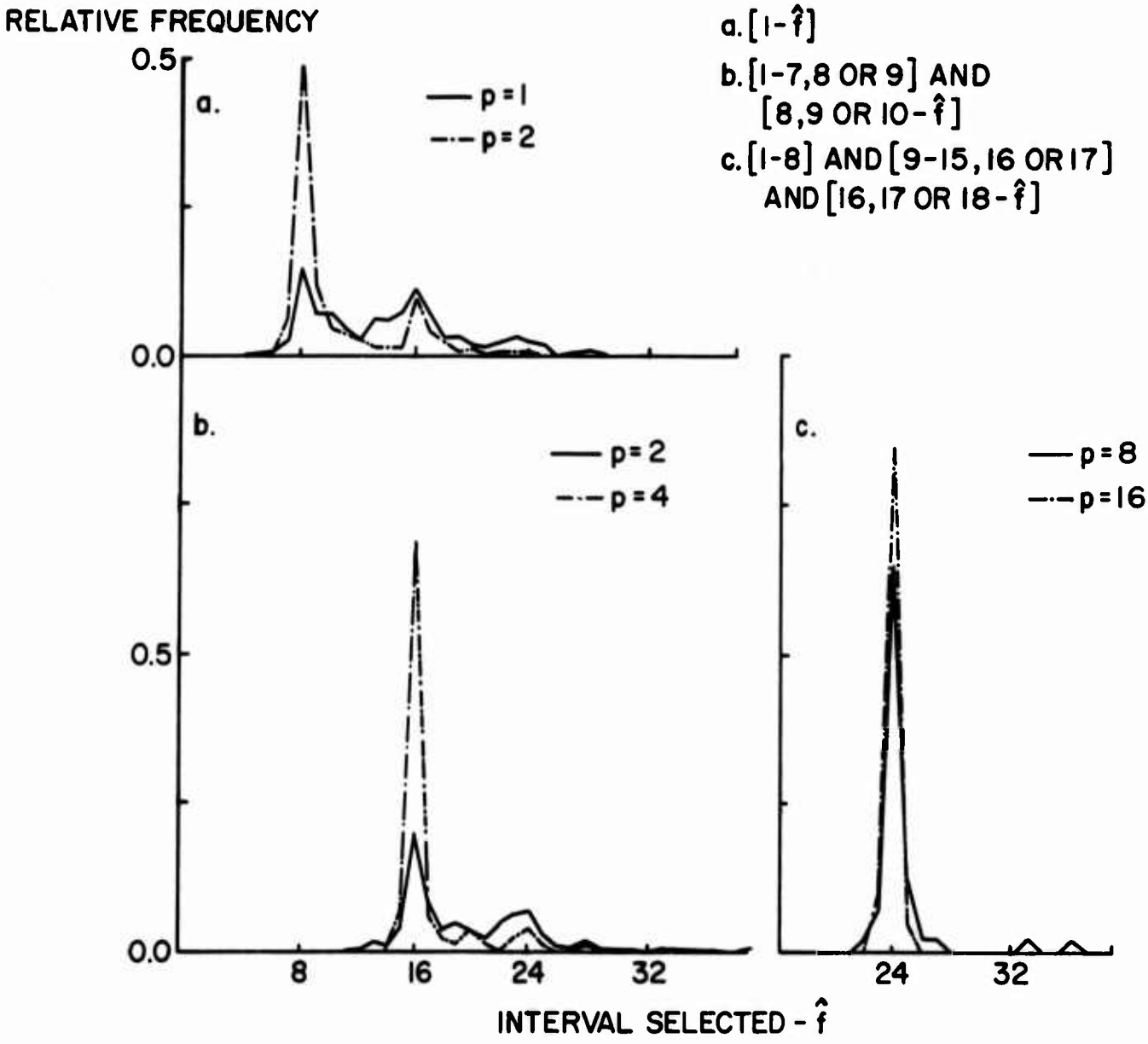
The results of the simulation are shown in Figure 6. We denote $\{(f, t) \mid f_1 \leq f \leq f_2, 1 \leq t \leq 16\}$ by $[f_1 - f_2]$. Ideally, our procedure should first separate [1-8] from [9-40], then separate [9-16] from [17-40], then separate [17-24] from [25-40] and finally separate [25-32] from [33-40]. Each of these divisions should occur with the lower-frequency side having the extra component. This ideal behavior occurred four times when p = 16.

In Figure 6a, the relative frequencies for the first step are shown for p = 1 and 2. In the first step, the maximum of $d_e(S)$ always occurred for an S of the form $\{(f, t) \mid 1 \leq f \leq \hat{f}, 1 \leq t \leq 16\}$. The possibilities division in time and the maximum occurring for the high-frequency side did not occur. For p = 2, the correct division was made in almost half the replications. For p = 1, the divisions were more spread out. For p = 4, the correct first division was always made with one exception.

In Figure 6b, we show for the second step the sum of the relative frequencies of the interval pairs $\{[1\text{-}7], [8\text{-}\hat{f}]\}$, $\{[1\text{-}8], [9\text{-}\hat{f}]\}$, and $\{[1\text{-}9], [10\text{-}\hat{f}]\}$. For p = 4, all replications had this form, one having been created by first selecting [1-16] and then dividing it into [1-8] and [9-16]. Note that when p = 4, [7, 8 or 9-16] occurred in over half the replications. For p = 2, 131 of the replications had the form appropriate to Figure 6b, 29 of these having been created by first selecting $[1\text{-}\hat{f}]$ and then dividing it.

In Figure 6c, we show for the third step the sum of the relative frequencies of the interval triples $\{[1\text{-}8], [9\text{-}15], [16\text{-}\hat{f}]\}$, $\{[1\text{-}8], [9\text{-}16], [17\text{-}\hat{f}]\}$, and

FIGURE 6. SIMULATION FOR FIVE SUBINTERVALS WITH DIFFERING SPECTRAL MATRICES



$\{[1\text{-}8], [9\text{-}17], [18\text{-}\hat{f}]\}$. For p = 4, the results were not even approximately correct more than half the time. For p = 8 and 16, the first division always separated [1-8] from [9-40]. For p = 8, one replication did not have the form appropriate to Figure 6c. For p = 16, they all did. Note that $\hat{f} = 24$ occurred over half the time for both p = 8 and 16. Even for p = 16, the results of the fourth step were usually incorrect.

A few generalizations about the performance of our procedure in the situation simulated seem noteworthy. We see that doubling p does not quite double the sensitivity and that when $u^* F_0^{-1} u$ (which equals u^*u when $p \geq 4$) is greater than one, the procedure performs reasonably. The results for p = 16 may indicate that increasing p may not increase sensitivity when the size of frequency-time interval is too small.

The computer time needed for the simulation is of interest. Each replication consisted of five cycles through the procedure (creating six intervals). On a CDC 6700, each replication took on the average 0.60 seconds for p = 1, 1.73 seconds for p = 2, 4.52 seconds for p = 4, 15.8 seconds for p = 8, and 71.7 seconds for p = 16.

REFERENCES

[1.] Akaike, Hirotugu, "Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models," Biometrika, 60, No. 2 (1973), 255-65.

[2.] Brillinger, David R., "The Canonical Analysis of Stationary Time Series," in P.R. Krishnaiah, ed., Multivariate Analysis-II, New York: Academic Press, 1969, 331-50.

[3.] Brillinger, David R., "The Frequency Analysis of Relations between Stationary Spatial Series," in Ronald Pyke, ed., Proceedings of the Twelfth Biennial Seminar of the Canadian Mathematical Congress, Montreal: Canadian Mathematical Congress, 1970, 39-81.

[4.] Carter, G. Clifford, Knapp, Charles H., and Nuttall, Albert H., "Estimation of the Magnitude-Squared Coherence Function Via Overlapped Fast Fourier Transform Processing," IEEE Transactions on Audio and Electroacoustics, AU-21, No. 4 (1973), 337-44.

[5.] Chernoff, Herman, Sequential Analysis and Optimal Design, Philadelphia: Society for Industrial and Applied Mathematics, 1972.

[6.] David, H.A., Order Statistics, New York: John Wiley and Sons, 1970.

[7.] Dempster, A.P., "An Overview of Multivariate Data Analysis," Journal of Multivariate Analysis, 1, No. 3 (1971), 316-46.

[8.] Durbin, J., "Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least-Squares Residuals," Biometrika, 56, No. 1 (1969), 1-15.

[9.] Fukunaga, Keinosuke, Introduction to Statistical Pattern Recognition, New York: Academic Press, 1972.

[10.] Goodman, N.R., "Statistical Analysis based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)," Annals of Mathematical Statistics, 34, No. 1 (1963), 152-77.

[11.] Goodman, N.R., and Dubman, M.R., "Theory of Time-Varying Spectral Analysis and Complex Wishart Matrix Processes," in P.R. Krishnaiah, ed., Multivariate Analysis-II, New York: Academic Press, 1969, 351-66.

[12.] Hannan, E.J., Multiple Time Series, New York: John Wiley and Sons, 1970.

[13.] Hauck, Walter W., Jr., and Portnoy, Stephen L., "Local Principal Components: A Method for Determining the Dimensionality of Non-Linear Data Structures," Research Report CP-8, Department of Statistics, Harvard University, 1972.

[14.] James, Alan T., "Distributions of Matrix Variates and Latent Roots Derived from Normal Samples," Annals of Mathematical Statistics, 35 (1964), 475-501.

[15.] Jenkins, Gwilym M. and Watts, Donald G., Spectral Analysis and Its Applications, San Francisco: Holden Day, 1968.

[16.] Khatri, C.G., "Distribution of the Largest or the Smallest Characteristic Root under Null Hypothesis Concerning Complex Multivariate Normal Populations," Annals of Mathematical Statistics, 35 (1964), 1807-10.

[17.] Khatri, C.G., "Classical Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution," Annals of Mathematical Statistics, 36 (1965), 98-114.

[18.] Koenig, W., Dunn, H.K., and Lacy, L.Y., "The Sound Spectrograph," Journal of the Acoustical Society of America, 18, No. 1 (1946), 19-49.

[19.] Krishnaiah, P.R., "On the Exact Distribution of the Statistics Based on the Eigenvalues of Complex Random Matrices," Preprint, 1973.

[20.] Liggett, Walter S., Jr., "Passive Sonar: Fitting Models to Multiple Time Series," Proceedings of the NATO Advanced Study Institute on Signal Processing, Loughborough, U.K., (1972). (To be published by Academic Press).

[21.] Loève, Michel, Probability Theory, Third Edition, Princeton: D. Van Nostrand Co., 1963.

[22.] Otnes, Robert K., and Enochson, Loren, Digital Time Series Analysis, New York: John Wiley and Sons, 1972.

[23.] Parzen, Emanuel, "Multiple Time Series Modeling," in P.R. Krishnaiah, ed., Multivariate Analysis-II, New York: Academic Press, 1969, 389-409.

[24.] Pillai, K.C. Sreedharan, and Kanta, Jayachandran, "Power Comparisons of Tests of Equality of Two Covariance Matrices Based on Four Criteria," Biometrika, 55, No. 2 (1968), 335-42.

[25.] Priestley, M.B., "Evolutionary Spectra and Non-stationary Processes," Journal of the Royal Statistical Society, Ser B, 27, No. 2 (1965), 204-37.

[26.] Priestley, M.B., and Rao, T. Subba, "A Test for Non-stationarity of Time Series," Journal of the Royal Statistical Society, Ser. B, 31, No. 1 (1969), 140-49.

[27.] Priestley, M.B., Rao, T. Subba, and Tong, H., "Identification of the Structure of Multivariable Stochastic Systems," Third International Symposium on Multivariate Analysis, Wright State University, Dayton, Ohio (June, 1972).

[28.] Rao, T. Subba, and Tong, H., "A Test for Time-dependence of Linear Open-loop Systems," Journal of the Royal Statistical Society, Ser B, 34, No. 2 (1972), 235-50.

[29.] Singleton, Richard C. and Poulter, Thomas C., "Spectral Analysis of the Call of the Male Killer Whale," IEEE Transactions on Audio and Electroacoustics, AU-15, No. 2 (1967), 104-13.

[30.] Tick, Leo J., "Estimation of Coherency," in Bernard Harris, ed., Spectral Analysis of Time Series, New York: John Wiley and Sons, 1967, 133-52.

[31.] Tukey, John W., "Data Analysis, Computation, and Mathematics," Quarterly of Applied Mathematics, 30, No. 1 (1972), 51-65.

[32.] Wahba, Grace, "On the Distribution of Some Statistics Useful in the Analysis of Jointly Stationary Time Series," Annals of Mathematical Statistics, 39, No. 6 (1968), 1849-62.